MAY 3 1982

830416

NAS 1.26:3547

ORIGINAL

COMPLETED

NASA Contractor Report 3547

Measurement of Speech Levels in the Presence of Time Varying Background Noise

Karl S. Pearsons and Richard D. Horonjeff

CONTRACT NAS1-16521
APRIL 1982

NASA

102

NASA Contractor Report 3547

Measurement of Speech Levels in the Presence of Time Varying Background Noise

Karl S. Pearsons and Richard D. Horonjeff
Bolt Beranek and Newman Inc.
Canoga Park, California

Prepared for
Langley Research Center
under Contract NAS1-16521

National Aeronautics
and Space Administration

**Scientific and Technical
Information Branch**

1982

Use of trade names or names of manufacturers in this report does not constitute an official endorsement of such products or manufacturers, either expressed or implied, by the National Aeronautics and Space Administration.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	A REVIEW OF SPEECH MEASURES	2
III.	APPROACH FOR DEVELOPMENT OF SHORT TERM SPEECH MEASURE	5
	A. Spectral Characteristics of Speech	5
	B. Temporal Characteristics of Speech	10
IV.	VALIDATION OF PROPOSED SPEECH MEASURES	12
	A. Vowel Equalization Filters, Phase I	12
V.	SPEECH MEASURES IN THE PRESENCE OF TIME VARYING BACKGROUND NOISE - PHASE II	21
	A. Subjects	21
	B. Procedure	22
	C. Analysis	27
	D. Results	28
VI.	DISCUSSION	41
VII.	CONCLUSIONS	45
	REFERENCES	47
	APPENDIX A - EQUIVALENT PEAK LEVEL (EPL)	49
	APPENDIX B - INSTRUMENTATION	55
	APPENDIX C - INSTRUCTIONS TO TEST SUBJECTS	65
	APPENDIX D - PREPARED TEXT	69

I. INTRODUCTION

Noise can undeniably disturb speech communication. The communication may be in the form of casual conversation, or lecturing, or it may be a command or warning of danger. As noise levels increase, people tend to raise their voices in an effort to make themselves heard over the intruding background noise. However, at some point people are no longer willing to raise their voice further and stop talking altogether.

Estimation of speech intelligibility in the presence of a competing background noise requires knowledge of both the background noise level and the speech level at the listener's ear. With this information, a measure known as articulation index (AI) may be calculated which in turn provides an estimate of the percentage of words or sentences correctly understood in a constant background noise environment. Because speech fluctuates considerably in level even for a constant vocal effort, previously used speech measures have been specified by an average of levels determined over a period of at least 10 seconds. Thus, for many events such as aircraft flyovers it is not possible to determine the speech level changes during the occurrence of the noise event. Research has been undertaken to determine the feasibility of a short-term measure of speech, or vocal effort, to allow investigations of the effect of short term transient events on

speech level. If successful, estimates of intelligibility could then be made for intrusions such as aircraft fly-overs or truck passbys.

This report is a summary of the research on the development of such a short term speech measure. Section II provides a review of the speech measures employed for long term speech measures. Section III provides a summary of the approach taken to develop a short term speech measure from a spectral and temporal standpoint. Section IV presents a brief investigation of a number of alternative speech measures. A test to validate a short term speech measure is described in Section V. A discussion of the suggested short term speech measure is provided in Section VI followed by general conclusions of investigations carried out in this research in Section VII.

II. A REVIEW OF SPEECH MEASURES

Speech measurements were originally developed for use in telephone or broadcast situations. Researchers in speech intelligibility have also employed various measures for speech levels. As a result, speech in currently measured using several different techniques. The techniques range from noting speech peaks on a meter to sophisticated averaging techniques using digital computers. One of the earlier measures of speech utilized a VU-(volume unit) meter. These meters were employed in broadcast studios or on tape recorders. The method

suffered from the multitude of nonstandard meters as well as the inability of the meter to "track" the speech levels. The vu-meter, properly calibrated, does provide a method of determining speech levels which will not overload recorders or broadcast modulators. For speech measurements, it is used to provide a convenient measure of repeating a speech level especially for a carrier phrase in an intelligibility test.

Graphic level recorders have also been used but variability between units even for the same settings produced different results. Further different definitions of what aspect of the speech pattern (e.g., peaks or average) make it difficult to compare results across studies.

Researchers have tried to define a simple measure for describing speech level in order to quantify the level for a normal speaking voice. The result is a measure termed "long term rms level". However different methods of measurement do not always produce the same numerical values.

An early study to determine the statistics of speech by Dunn and White (ref. 1) indicates speech levels cover a dynamic range of about 36 dB. Later Beranek (ref. 2) in a reanalysis of the Dunn and White data suggested a 30 dB range if pauses in the speech were removed. This range is used today in determining articulation index (AI) as defined by an ANSI standard (ref. 3). Several

other researchers including Benson & Hirsh (ref. 4) and French & Steinberg (ref. 5) have used the Dunn and White data (ref. 1), but no other study on statistics of speech levels exists in published form at this time.

The method suggested by the AI standard as an approximation to long term RMS is an average of the maximum speech peaks determined on a sound level meter - slow - reduced by 3 dB. Since the difference between A-weighted level and overall or C-weighted level for "normal" speech is 3 dB, then the same quantity may be determined by measuring directly with the A-weighting on a slow sound level meter. However this value should be reported as overall sound pressure level or long term RMS speech level rather than A-weighted level.

Because of the continuing trend to use A-level for measuring environmental noise, A-level has also been used to measure speech level (ref. 6). However one should again remember that overall sound pressure level is 3 dB higher than A-weighted sound pressure level for normal voice level. As voice level becomes greater, the difference between overall and A-weighted sound level is smaller until for a shout there is no difference at all between overall sound pressure level and A-weighted level.

A relatively new measure of speech has been developed at Bell Telephone Laboratories (ref. 7), which relates primarily to the peaks in speech level. It is termed equivalent peak level (EPL). One of its primary advant-

MICROFILMED FROM
BEST AVAILABLE COPY

ages is that it only accepts measurements while speech is being uttered. Thus it accounts for pauses in speech rather than averaging them in with the speech level. Again, the measure was intended to be used for continuous discourse or a minimum of 5 to 10 seconds. Further details on the measure are included in Appendix A.

A recent comparison of various measures of speech has been made by Steeneken & Houtgast (ref. 8). The study indicated that some of the measures were more reproducible than others, however, most, except for the peak reading measures, fell within an accuracy of 1 dB. Methods using the A-weighting of the sound level meter provided a closer relationship with intelligibility ratings than methods without any band pass limiting. The authors selected several measures which seem preferred over other measures under test for speech level quantification. The results also provide an indication of the mean difference in levels between the various measures under task.

III. APPROACH FOR DEVELOPMENT OF SHORT TERM SPEECH MEASURE

A. Spectral Characteristics of Speech

Although many methods exist for measuring speech level for a long period of time (e.g., longer than 10 seconds), no method exists to make reliable speech measurements for short periods of time (e.g., a 1 to 2 second period).

MICROFILMED FROM
BEST AVAILABLE COPY

This is partially due to the difference in level of various speech phonemes and partly due to the random grouping of speech sounds in language. Vowels account for the majority of energy in speech sounds and tend to exhibit a longer duration than consonants. However, all vowels are not of the same speech level even when spoken with the same vocal effort. The first step in the development of short-term speech measure was to design a "vowel equalization network" to compensate for the unequal speech levels produced by different vowels. Unfortunately, data on the relative levels of vowels was not consistent. Further, it appeared that relative vowel levels were dependent on vocal effort as shown in Figure 1. One source of data (ref. 9) on vowel measurements taken at normal voice levels suggests that vowels increase in level monotonically at a rate of 7 dB per decade of frequency. The result is a 3 dB increase over the vowel frequency range from 250 to 750 Hz. Although the effect seemed small, a filter was designed to compensate for this difference. The frequency response of this filter is shown in Figure 2 (V2). Another source (ref. 6) which analyzed speech levels of vocal efforts ranging from normal to shout as shown in Figure 1 suggested a more complicated compensating filter indicated as V1 in Figure 2. Also shown in Figure 2 is the response for a band limiting filter which encompasses vowel sounds contributing to overall sound pressure level without any vowel equalization. For comparison the response for the A-weighting is also indicated in Figure 2. A numerical tabulation of the one-third octave band weightings for each of these filters is shown in Table I.

MICROFILMED FROM
BEST AVAILABLE COPY

88.3
87.8

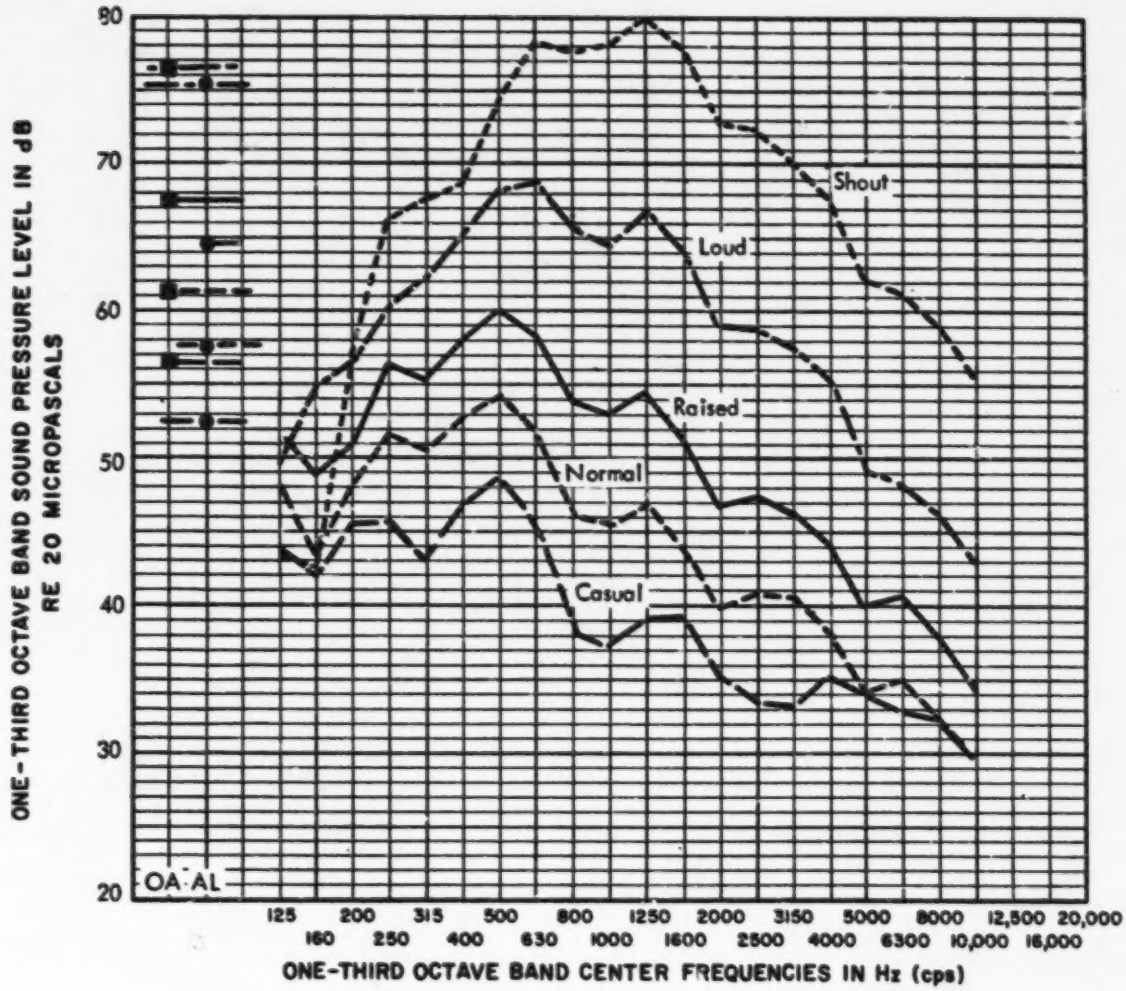


FIGURE 1. AVERAGE SPEECH SPECTRA FOR MALES AT FIVE VOCAL EFFORTS (PEARSONS & BENNETT, ref. 6)

MICROFILMED FROM
BEST AVAILABLE COPY

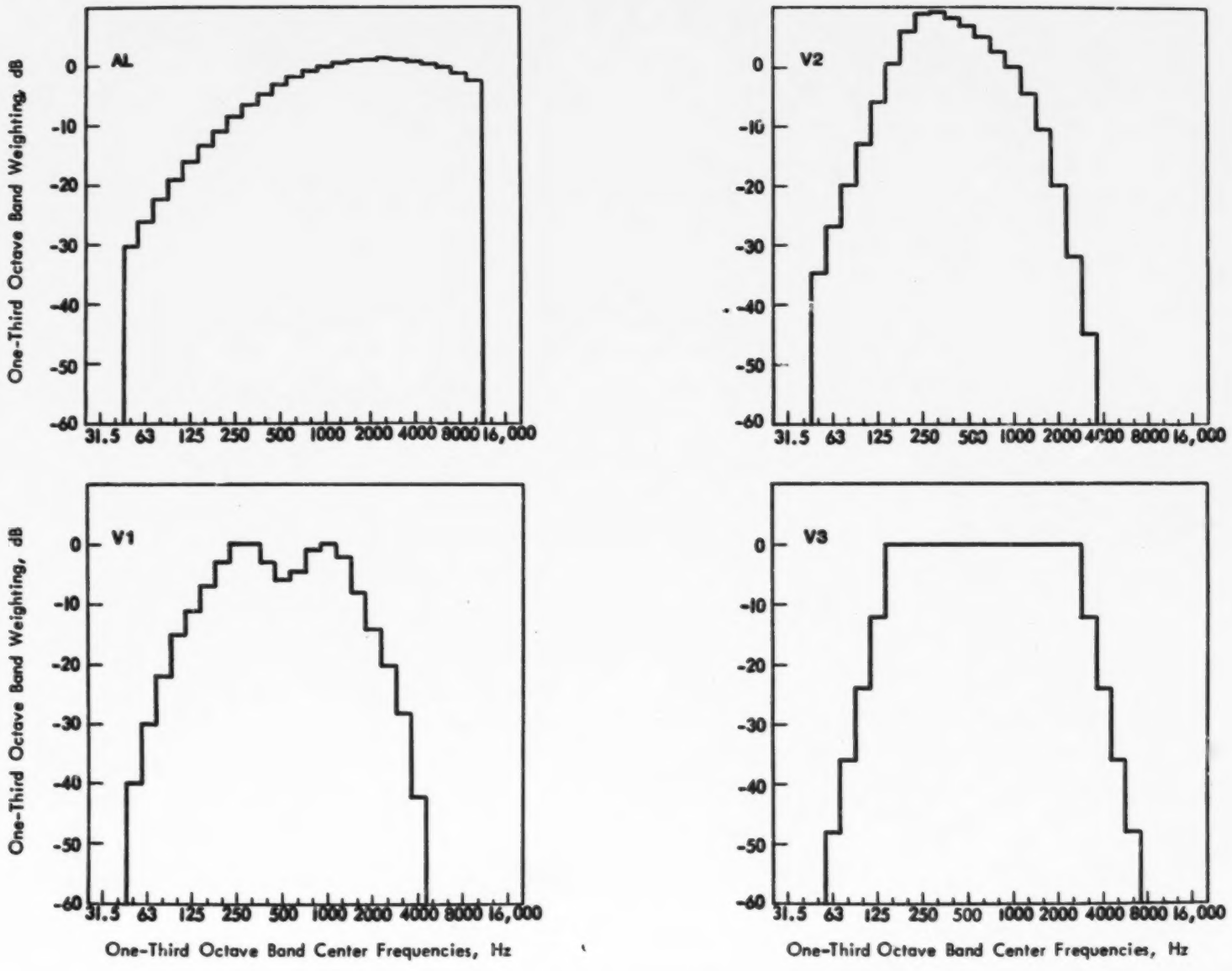


FIGURE 2. ONE-THIRD OCTAVE BAND FREQUENCY WEIGHTINGS FOR VARIOUS SPEECH LEVEL MEASURES

BEST AVAILABLE COPY
MICROFILMED FROM

TABLE I. ONE-THIRD OCTAVE BAND WEIGHTINGS FOR VARIOUS MEASURES OF VOCAL EFFORT

1/3 OB Center Frequency (Hertz)	SPEECH MEASURE (dB)			
	A-LEV	V1	V2	V3
50	-30.2	-40.0	-35.0	-60.0
63	-26.2	-30.0	-27.0	-48.0
80	-22.5	-22.0	-20.0	-36.0
100	-19.1	-15.0	-13.0	-24.0
125	-16.1	-11.0	- 6.0	-12.0
160	-13.4	- 7.0	0.5	0.0
200	-10.9	- 3.0	6.0	0.0
250	- 8.6	0.0	9.0	0.0
315	- 6.6	0.0	9.2	0.0
400	- 4.8	- 3.0	8.3	0.0
500	- 3.2	- 6.0	7.0	0.0
630	- 1.9	- 4.5	5.0	0.0
800	- 0.8	- 1.0	2.5	0.0
1000	0.0	0.0	0.0	0.0
1250	0.6	- 2.0	- 4.5	0.0
1600	1.0	- 8.0	-10.5	0.0
2000	1.2	-14.0	-20.0	0.0
2500	1.3	-20.0	-32.0	0.0
3150	1.2	-28.0	-45.0	-12.0
4000	1.0	-42.0	-60.0	-24.0
5000	0.5	-60.0	-60.0	-36.0
6300	- 0.1	-60.0	-60.0	-48.0
8000	- 1.1	-60.0	-60.0	-60.0
10000	- 2.5	-60.0	-60.0	-60.0

MICROFILMED FROM
BEST AVAILABLE COPY

B. Temporal Characteristics of Speech

Independent of the spectral equalization issue is that of an appropriate short term integration time over which a discrete estimate of the speech level may be made. As cited earlier, fluctuations in speech level are quite sizeable over short periods of time, even when the speaker is reciting continuous, prepared text. To put these fluctuations in perspective, consider the speech samples represented in Figure 1 for various vocal efforts of speech. With a one-second integration period, discrete samples of the overall rms sound pressure level exhibited a standard deviation of the order of one to two decibels. In contrast, a Gaussian source of equivalent spectral content has a standard deviation on only 0.1 to 0.2 decibels. Thus, with speech we are dealing with a source which is an order of magnitude greater in variability than a Gaussian one (and would thus require one-hundred times longer to estimate a mean value to the same level of precision as that of the Gaussian source). Clearly, the longer the averaging time, the more stable the estimated speech level; but lengthy averaging times (such as the 10 seconds mentioned in the previous subsection) are not suitable if an attempt is to be made at tracking vocal effort with changing background noise levels (particularly short term intrusions, such as those produced by passing aircraft). In 10 seconds the majority of an aircraft intrusion could easily have come and gone. The problem, then, is to decide how much change in background level one is willing to tolerate during a discrete

MICROFILMED FROM
BEST AVAILABLE COPY

sample integration of the speech, and then to examine typical rise/decay rates of background noise intrusions to make an intelligent selection of this integration time.

Starting with the speech level itself, suppose the following baseline criterion were established: an upward or downward change in background noise level should not produce an expected change in long term rms speech level of more than one decibel during the integration period. An estimate of the allowable change in background noise level may be made by referring to the review by Lane & Tranel (ref. 10) and the work of Heusden, Plomp and Pols (ref. 11). Rather than a one-for-one relationship, these studies suggests that speech levels change by only about 0.3 dB for every decibel change in background noise. Conversely, background noise must change approximately 3 dB to evoke a one decibel change in speech level.

The upper limit on the integration period may be determined by considering the rise/decay rates of the expected intrusions to determine how quickly this 3 dB change is likely to occur. An aircraft flyover, for example, can produce a vast range of rise/decay rates depending on its speed and distance from the observer. If the aircraft is close enough to create an indoor speech interference problem, however, the distances are likely to be short (on the order of only a few thousand feet) and the distance/speed combinations could produce A-weighted 10 dB down durations as short as 6 to 10 seconds. Assuming a triangular time pattern, these durations cor-

MICROFILMED FROM
BEST AVAILABLE COPY

respond to 3.3 and 2.0 dB per second rise/decay rates. Dividing the permissible 3 dB change in background level by these rates yields optimal integration times of 0.9 to 1.5 seconds (nominally 1 to 2 seconds). This integration time range is optimal in the sense that it can be expected to provide the most favorable temporal window for observing short term changes in speech level. Lesser integration times increase the variability of the speech measurement while doing nothing to materially improve the ability to track a changing vocal effort. And greater integration times sacrifice the ability to track a changing vocal effort by simply averaging over too long a period of time.

IV. VALIDATION OF PROPOSED SPEECH MEASURES

A. Vowel Equalization Filters, Phase I

To determine which of the vowel equalization methods depicted in Figure 2 produced measurements with the least variation, speech samples were obtained from recordings made for an earlier project investigating levels of speech for normal, raised, loud, and shouting vocal efforts (ref. 6). The phrases employed in the test were:

> "Joe took my father's shoe bench out. She was waiting at my lawn".

These phrases were used since they contain all sounds normally found in the English language.

MICROFILMED FROM
BEST AVAILABLE COPY

In addition, recordings for the same project were made during "casual" conversations with the subject prior to the test proper. The speech material was analyzed for a minimum of 10 seconds using data from 6 subjects, three males and three females ranging in age from 16 to 45 years.

Sound levels were determined every second by reading an "RMS slow" one-third octave band spectrum and calculating the various spectrally weighted measures. The equivalent peak level was determined by making a separate pass through the magnetic tape. Means and standard deviations were calculated over the length of the speech sample for each of the 6 speech measures (overall sound pressure level, A-level, V1, V2, V3 weighted sound level and equivalent peak level) and the 6 subjects. Results are shown in Table II for individual subjects.

Comparisons were then made between each of the measures and the overall sound pressure level as shown in Figures 3 through 7. Each figure (representing individual measures) contains two graphs. One graph shows the observed relationship between mean values of the speech, the other shows a similar relationship between the standard deviations. Mean values (the left hand graph) of all measures show a strong linear relationship with overall sound pressure level, a not altogether unexpected observation. The finding is comforting, however, since adoption of any of the measures as a short term speech metric would still allow direct comparison with any prior speech levels reported in the literature.

MICROFILMED FROM
BEST AVAILABLE COPY

MICROFILMED FROM
BEST AVAILABLE COPY

TABLE II. VARIOUS MEASURES OF SPEECH LEVEL

Subject	Measure	SPONTANEOUS CONVERSATION		REPEATED PHRASES: NORMAL		RAISED		LOUD		SHOUT	
		$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ
3	OA	55.1	1.68	62.4	0.70	67.2	0.64	78.3	1.02	88.9	1.44
	AL	48.5	1.50	57.2	0.90	63.1	0.77	76.2	1.19	88.1	1.56
	V1	52.0	1.87	59.5	0.88	64.7	0.70	75.5	1.11	86.7	1.67
	V2	61.7	1.82	69.4	0.98	74.3	0.67	84.1	0.88	92.7	1.12
	V3	54.0	1.62	61.9	0.81	66.9	0.66	78.3	1.02	88.9	1.44
	EPL	64.7	1.84	69.4	1.97	74.8	1.10	86.6	1.84	97.4	2.29
21	OA	51.7	3.01	58.3	1.93	69.7	0.80	78.2	1.62	88.3	3.05
	AL	47.9	3.56	56.0	2.09	68.7	0.90	78.3	1.75	88.7	2.96
	V1	48.2	2.98	55.2	1.72	66.4	0.91	74.3	1.83	84.2	3.15
	V2	57.7	3.16	64.4	1.88	74.4	0.93	79.7	1.10	87.5	4.00
	V3	51.5	3.04	58.2	1.93	69.5	0.82	77.9	1.70	88.1	3.10
	EPL	61.6	1.91	66.7	2.43	78.3	0.97	87.4	2.42	95.9	2.39
33	OA	57.0	3.42	58.5	0.84	63.4	0.84	66.7	0.66	78.2	1.38
	AL	52.8	3.62	55.6	0.93	61.1	1.06	65.1	0.75	78.4	1.52
	V1	54.4	3.77	55.1	0.93	60.4	0.88	64.0	0.67	74.6	1.34
	V2	63.9	3.87	64.9	0.91	69.5	0.81	72.2	0.67	79.4	0.81
	V3	56.8	3.42	58.2	0.87	63.2	0.80	66.6	0.66	78.1	1.42
	EPL	63.9	4.41	66.2	2.56	71.8	1.38	75.1	0.70	87.1	3.05
56	OA	53.0	3.43	61.3	1.34	64.7	0.66	74.9	0.90	85.3	1.04
	AL	48.0	3.63	57.3	1.65	62.4	0.87	75.1	0.89	86.1	1.10
	V1	50.7	3.10	59.2	1.15	61.6	0.75	69.8	1.22	77.9	1.09
	V2	60.2	3.35	68.9	1.23	71.2	0.70	75.5	1.41	82.6	0.97
	V3	52.9	3.33	61.2	1.36	64.7	0.65	74.2	0.87	84.2	0.94
	EPL	62.6	1.98	67.8	2.16	73.3	2.06	84.5	2.19	93.7	1.33
84	OA	57.8	1.96	62.1	1.26	67.2	1.14	74.8	2.43	86.7	1.77
	AL	50.9	1.57	57.0	1.66	64.4	1.12	72.7	2.65	86.0	1.93
	V1	53.8	2.02	58.9	1.42	64.7	1.01	71.8	2.51	83.2	1.91
	V2	63.0	2.13	68.5	1.36	74.0	1.19	80.6	2.16	90.7	1.64
	V3	56.2	2.01	61.2	1.36	67.0	1.10	74.6	2.42	86.5	1.76
	EPL	64.5	3.69	69.7	1.96	76.1	1.81	84.3	1.35	95.5	2.68
93	OA	60.4	4.40	61.3	0.86	67.5	0.54	73.0	0.66	85.0	1.15
	AL	53.6	5.71	54.8	1.11	63.0	0.70	69.8	0.66	83.7	1.12
	V1	56.0	4.32	58.3	0.93	64.6	0.54	69.6	0.71	82.4	1.22
	V2	65.6	4.52	68.1	0.91	74.2	0.50	79.2	0.67	89.7	1.38
	V3	58.2	5.36	59.8	0.99	66.8	0.56	72.8	0.67	84.9	1.16
	EPL	65.4	3.70	69.6	2.25	76.1	1.28	81.6	1.85	93.9	1.97

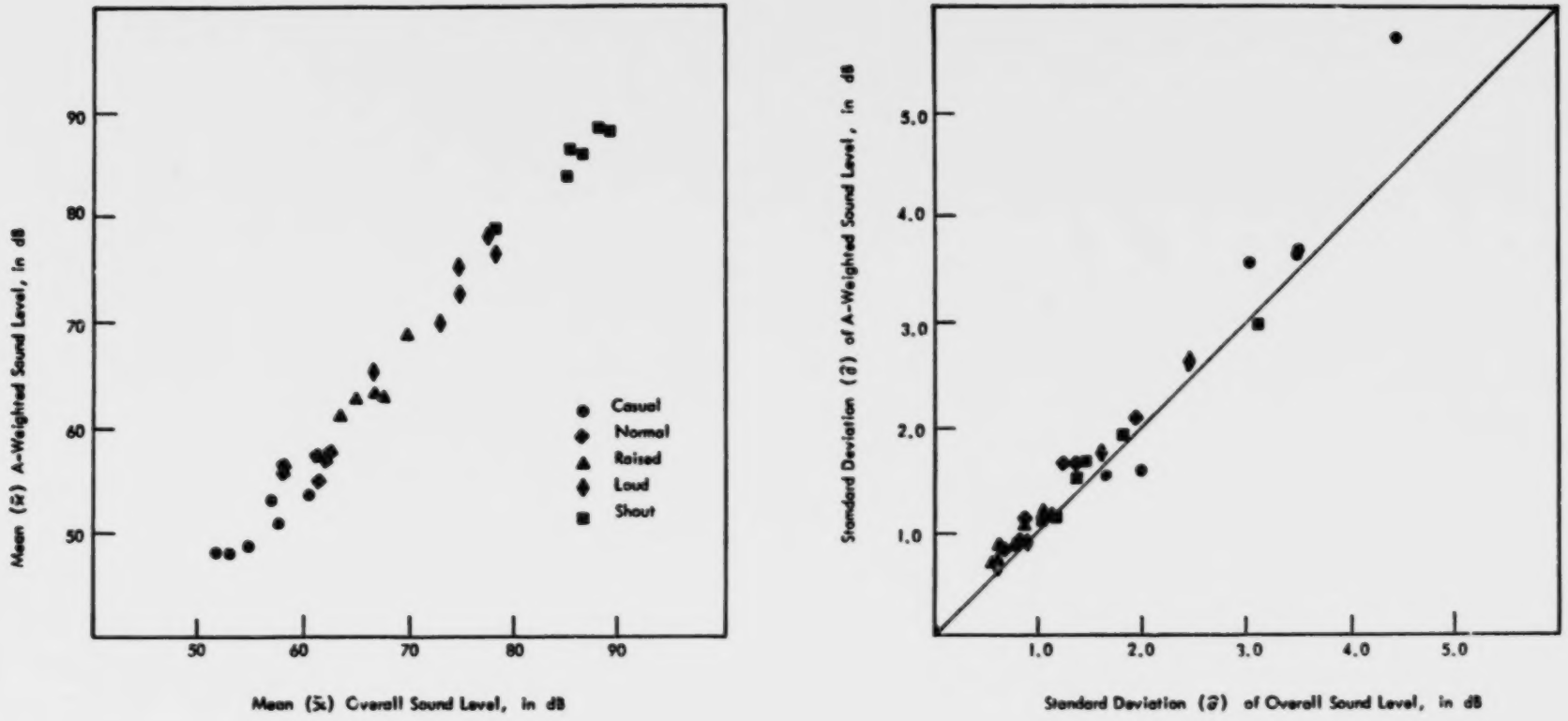


FIGURE 3. COMPARISON OF MEANS AND STANDARD DEVIATIONS BETWEEN A-WEIGHTED AND OVERALL SOUND LEVELS

MICROFILMED FROM
BEST AVAILABLE COPY

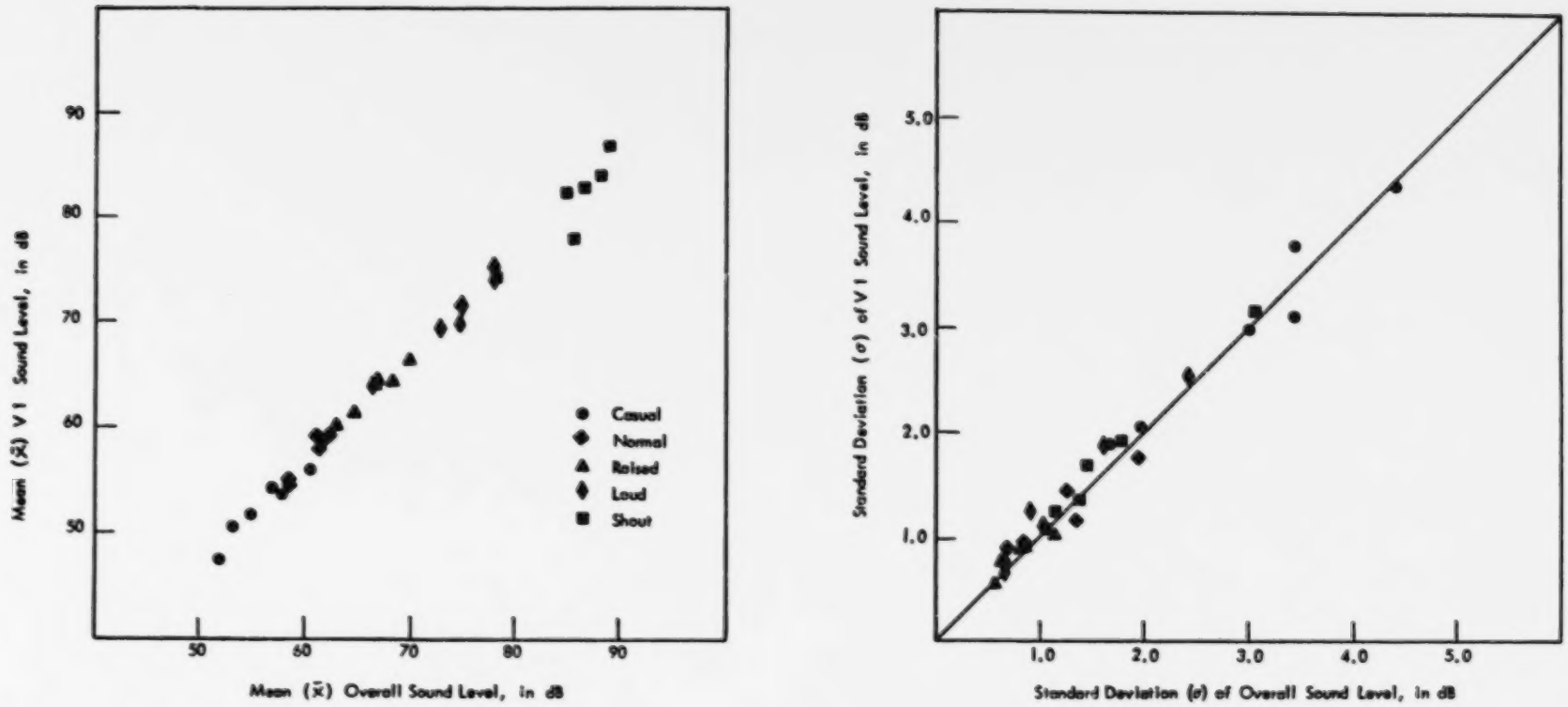
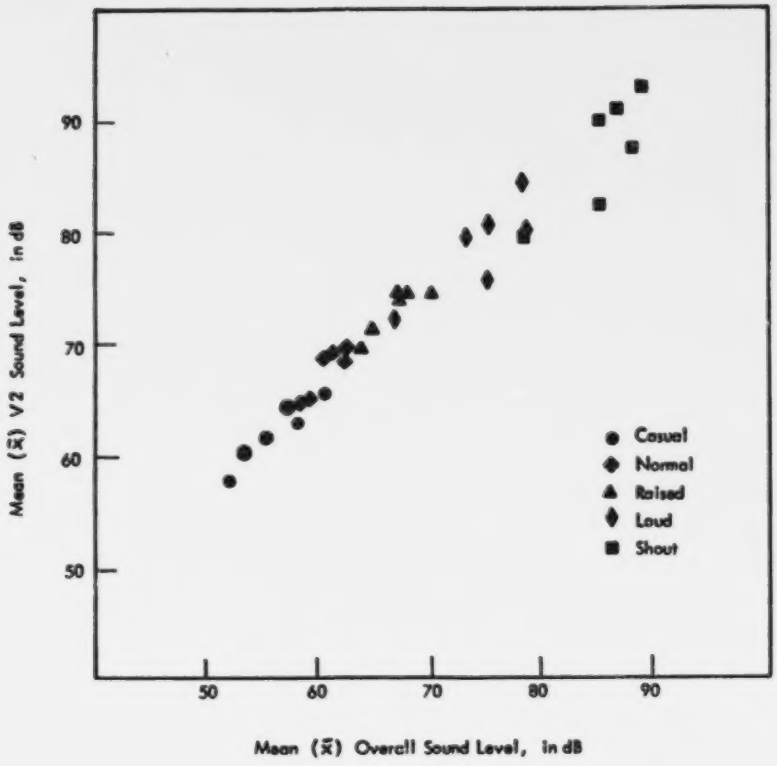


FIGURE 4. COMPARISON OF MEANS AND STANDARD DEVIATIONS BETWEEN V1 AND OVERALL SOUND LEVELS

BEST AVAILABLE COPY



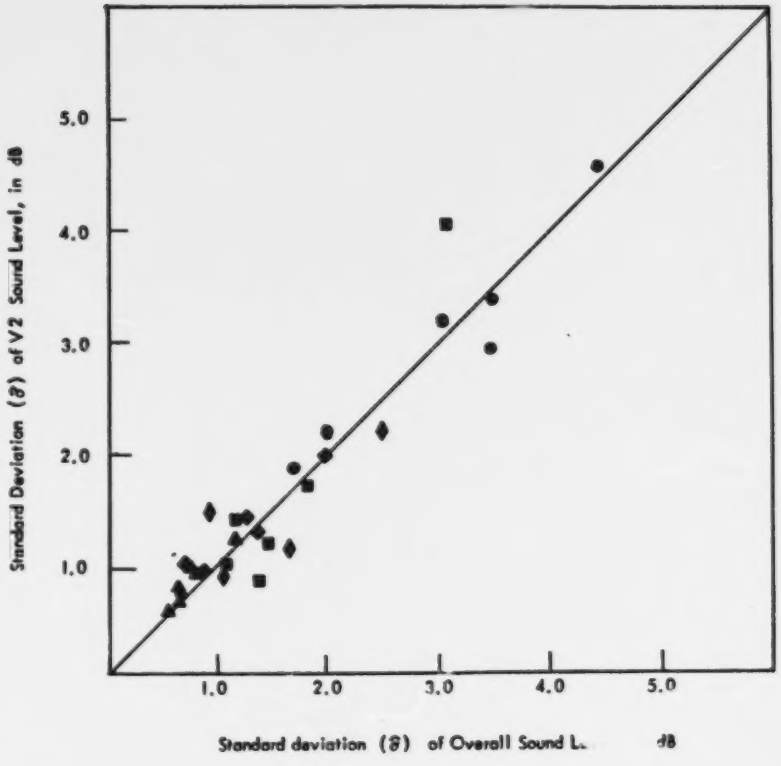


FIGURE 5. COMPARISON OF MEANS AND STANDARD DEVIATIONS BETWEEN V2 AND OVERALL SOUND LEVELS

MICROFILMED FROM
BEST AVAILABLE COPY

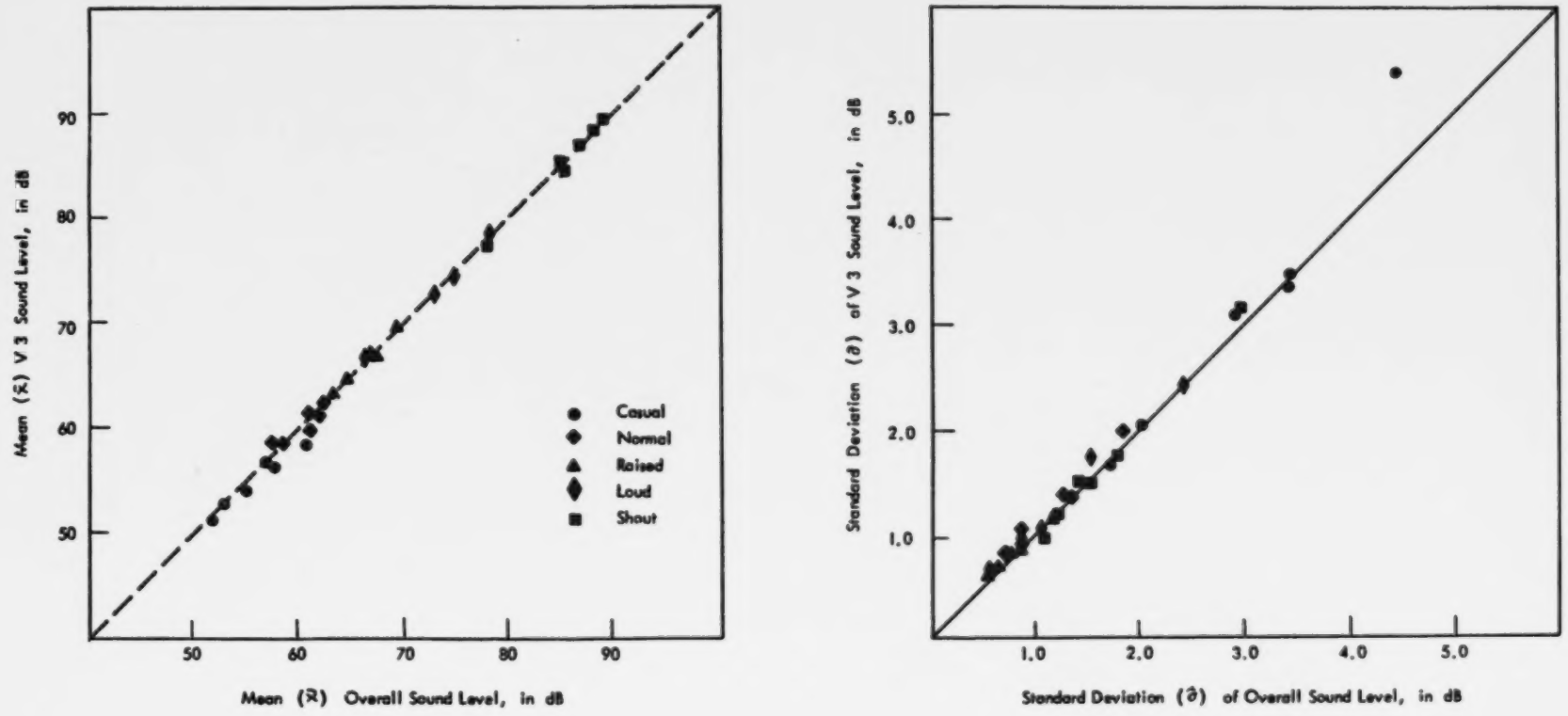


FIGURE 6. COMPARISON OF MEANS AND STANDARD DEVIATIONS BETWEEN V3 AND OVERALL SOUND LEVELS

MICROFILMED FROM
BEST AVAILABLE COPY

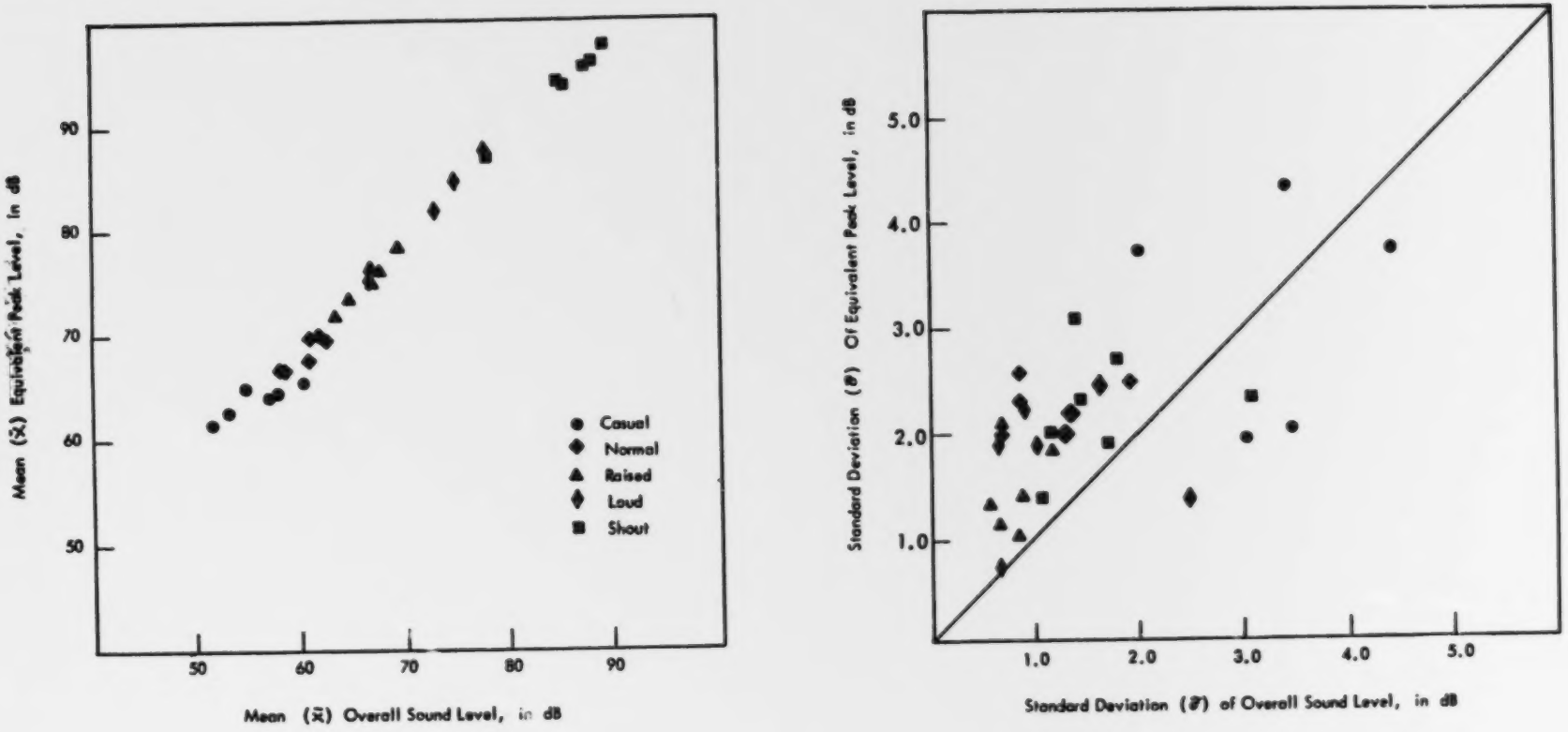


FIGURE 7. COMPARISON OF MEANS AND STANDARD DEVIATIONS BETWEEN EQUIVALENT PEAK LEVEL AND OVERALL SOUND LEVEL

MICROFILMED FROM
BEST AVAILABLE COPY

The standard deviation (right hand graph) shows how the moment to moment variability observed with any of the measures compared with that observed for the overall sound level. The diagonal line across the graph aids in comparing the relative variability of the two measures. If the data points tend to lie above the line then the alternative measure has more variability (i.e., is probably a poorer choice for a short term speech measure) than the overall sound level. Conversely, if the points tend to lie below the line then the alternative measure exhibits less variability than the overall sound level and would serve as a better short term measure. The alternative measure with the least variability is the one whose data points tend to lie the lowest with respect to the diagonal, and if all alternative measures tend to lie above the diagonal then the overall sound level has the least observed variability.

A review of Figures 3 through 7 quickly reveals that none of the alternative measures performs systematically better or worse than the overall sound level, in fact they generally do about the same with little exception. The equivalent peak level (Figure 7) exhibits more scatter in its performance, sometimes performing considerably better than the overall sound level, but also doing considerably worse on many occasions. For this reason the EPL is probably not the best choice for a short term measure without some refinement. Since the remaining measures appeared comparable in variability,

MICROFILMED FROM
BEST AVAILABLE COPY

the V3 weighted sound level was selected for use in the analysis of speech levels for the remainder of the project. This measure was chosen both for its simplicity as well as its potential of providing some limited noise reduction in noisy background level situations through band limiting. It also has the unique advantage (indicated in Figure 6) of having the same numerical value as overall SPL to facilitate comparison with other studies.

V. SPEECH MEASURES IN THE PRESENCE OF TIME VARYING BACKGROUND NOISE - PHASE II

The purpose of this phase was to apply optimal spectral weightings and temporal averaging techniques to recordings of continuous discourse speech recited in the presence of a time varying background noise. The question to be answered is whether or not the measured speech level tracks the changing background level, and if so how well. To answer this question a pair of synchronized level records are required, one for the speech and one for the background noise. From these records both visual assessments as well as statistical comparisons (cross-correlations, etc.) may be made.

A. Subjects

Twelve audiometrically screened subjects (2 males and 10 females) participated in this phase of the study. All subject's hearing levels were within 10 dB of normal

MICROFILMED FROM
BEST AVAILABLE COPY

hearing (ref. 12). Median age for the group was 21 years, ranging from 18 to 23. Subjects were paid to recite prepared text while simultaneously listening to prerecorded background noise over a binaural head set.

B. Procedure

The data acquisition methodology was influenced for the most part by inherent measurement constraints. Measurement of either speech or background noise in the presence of the other presents an immediate signal-to-noise ratio problem, how to exclude one to obtain a faithful measurement of the other. The problem is particularly accentuated when the background noise is of sufficient intensity to evoke an elevated vocal effort.

Measurement of the background noise by itself is relatively straightforward. If the background noise is prerecorded and suitable calibration procedures are used to ensure that all subjects are presented with the noise at the same level, these measurements can be made in the complete absence of the subject. A means for later time synchronization with the measured speech levels is of course required (this point is discussed in greater detail in subsequent paragraphs).

Measurement of the speech is a somewhat more complex issue since the background noise must be playing while the speech is being recorded. With the background noise playing through a loudspeaker, the signal-to-noise ratio (S/N) can be improved by moving the measuring microphone to within

MICROFILMED FROM
BEST AVAILABLE COPY

centimeters of the speaker's lips, thus increasing the speech signal level considerably over the more traditional one-meter measurement. The drawback of this technique is that microphone placement becomes extremely critical. Placement errors of one or two centimeters would result in one to three decibel differences in the measured sound level instead of a few tenths of a decibel for the same placement error at one meter. But even at a 5 cm distance an S/N much greater than 5 dB could not reasonably be expected.

An alternate approach is to play the background noise over a headset instead of a loudspeaker and to make speech level measurements with a conventional microphone at one meter. The major constraint in this approach is that the subject must receive unattenuated feedback of his own voice while he or she is listening to the background noise. This can be accomplished by the use of headsets which do not seal around the ear. Several commercially available headsets fulfill this requirement; the cushions are made of open cellular foam material with negligible loss below 4 kHz. Determination of sound levels actually heard by subjects may, admittedly, be prone to some small amount of error due to differences in fit and adjustment on individual subjects, but this limitation is more than offset by the convenience of a one-meter speech microphone placement with excellent S/N. Furthermore, only relative changes in speech and background levels are of primary interest; absolute levels are of lesser concern.

MICROFILMED FROM
BEST AVAILABLE COPY

An auxiliary solution to quantifying vocal effort is to measure epidermal vibration levels on the neck by a vibration transducer. The transducer responds to throat vibration but is insensitive to the background noise sound field. In preliminary tests, such a transducer exhibited signal-to-noise ratios over the entire frequency range of speech in excess of 30 dB, even with the speaker talking at low to moderate levels in the presence of a 90 dB(A) ambient noise field. This transducer is commonly referred to as a "throat microphone" and is used primarily to aid speech communication in noisy environments. It consists of an elastic band with two transducers (each the size of a nickel) which is clipped snugly around the neck. Because of the promise of this device for use outside the laboratory (where the source of intruding noise is not limited to headsets) it also was incorporated as a part of this experiment.

Appendix B presents a detailed description of instrumentation and recording techniques used to acquire the speech data. Briefly, however, each subject was seated alone in an anechoic chamber (of 2.4 x 3.0 x 2.3 meters interior dimensions) with a condenser microphone located one meter from the subjects' lips to record the speech. A set of written instructions (see Appendix C) was provided and the subject encouraged to ask questions of the experimenter. The subject was then given a set of foam cushion earphones and asked to adjust them for a comfortable fit. Next, one set of prepared text (see Appendix D) was issued and the subject asked to make him or herself familiar with the

MICROFILMED FROM
BEST AVAILABLE COPY

content (to minimize stumbling or other unintentional pauses during the recorded rescitation). The subject was instructed via intercom to start reading the prepared test as though he or she wished to communicate with someone one meter away. The background noise tape (reproduced by the earphones) as well as a separate tape transport for recording the speech were then started.

The background noise tape contained nine noise intrusions of nominal 10 to 20 second duration separated by 15 to 20 seconds of dead time between them. The intrusions consisted of three different signals (a steady state shaped Gaussian noise of 10 seconds duration, a triangular temporal pattern of shaped Gaussian noise with a 7 second 10 dB down duration, and a recorded aircraft flyover of 10 second 10 dB down duration). Each intrusion was presented at three different levels (nominally 65, 75 and 85 dB[A]). There was no other sound recorded on the tape. The intrusions were presented in random order with no signal following itself. Table III shows the presentation order. The total duration of the tape was 4.7 minutes and all subjects heard the same tape. A 1000 Hz sinusoid was recorded at the beginning of the tape (not heard by subjects) to provide a calibration voltage across the headset terminals.

Prerecorded on a second channel of this tape were two brief tone bursts, one preceding the first intrusion by about 15 seconds and the second trailing the last intrusion by the same amount. During playback to the test subjects

MICROFILMED FROM
BEST AVAILABLE COPY

TABLE III

SIGNAL PRESENTATION ORDER

Order	Signal	Maximum A-Level, dB
1	Steady state, shaped Gaussian noise	77
2	Recorded aircraft flyover	77
3	Steady state, shaped Gaussian noise	67
4	Recorded aircraft flyover	87
5	Time varying, shaped Gaussian noise	76
6	Recorded aircraft flyover	67
7	Time varying, shaped Gaussian noise	67
8	Steady state, shaped Gaussian noise	87
9	Time varying, shaped Gaussian noise	86

MICROFILMED FROM
BEST AVAILABLE COPY

these bursts were re-recorded on a separate channel of the speech tape recorder to enable subsequent synchronization of speech and background records.

The subject was told to continue reading the prepared text until verbally instructed to stop (after the second tone burst). In the event the subject finished the prepared text early he or she was instructed to start over at the beginning of the passage without pausing. Once finished the subject was issued a second set of text, the background noise tape was rewound, and the process repeated. Thus, when the recording session was complete each subject had read through 18 noise intrusions.

C. Analysis

Both background and speech signals were analyzed through a one-third octave band real time spectrum analyzer and digital computer. The equipment employed is described in detail in Appendix B. The spectrum analyzer performs continuous averaging in 24 independent one-third octave bands with a nominal RC time constant of one second and conforms to precision sound level meter "RMS SLOW" specifications. The computer digitized the sound level at a rate of one spectrum per second and stored this information in memory. The reference zero time for all analyses was always the first of the two synchronizing tone bursts. The second burst was used to ensure that no timing errors occurred, thus enabling any two records to be intercompared at any point in time. Two second average readings

MICROFILMED FROM
BEST AVAILABLE COPY

were computed digitally by energy averaging successive pairs of readings in each one-third octave band. The computer reported the one-third octave band spectrum time history, computed the overall, A-level and V3-level for these spectra and reported these time histories as well.

Background noise levels were determined in an anechoic chamber by instrumenting a lifesize, molded rubber human head with a one-half inch condenser microphone and microphone preamplifier in the ear cavity. The output of the preamplifier was connected directly to the real time spectrum analyzer. The headset used by the subjects was centered over the ears and voltage levels across the headset terminals brought into calibration. The background noise tape was then played and the levels digitized by the computer.

Speech levels were determined by playing the recorded-data tapes into the real time spectrum analyzer. A pistonphone calibration applied to the microphone and recorded at the beginning of each subjects' data served as an absolute reference.

From the digitized level records spectral and temporal relationships were plotted and correlations between speech and background levels performed. The following subsection describes the results in detail.

D. Results

Twelve subjects participated in this experiment. Both visual observations as well as statistical correlations

MICROFILMED FROM
BEST AVAILABLE

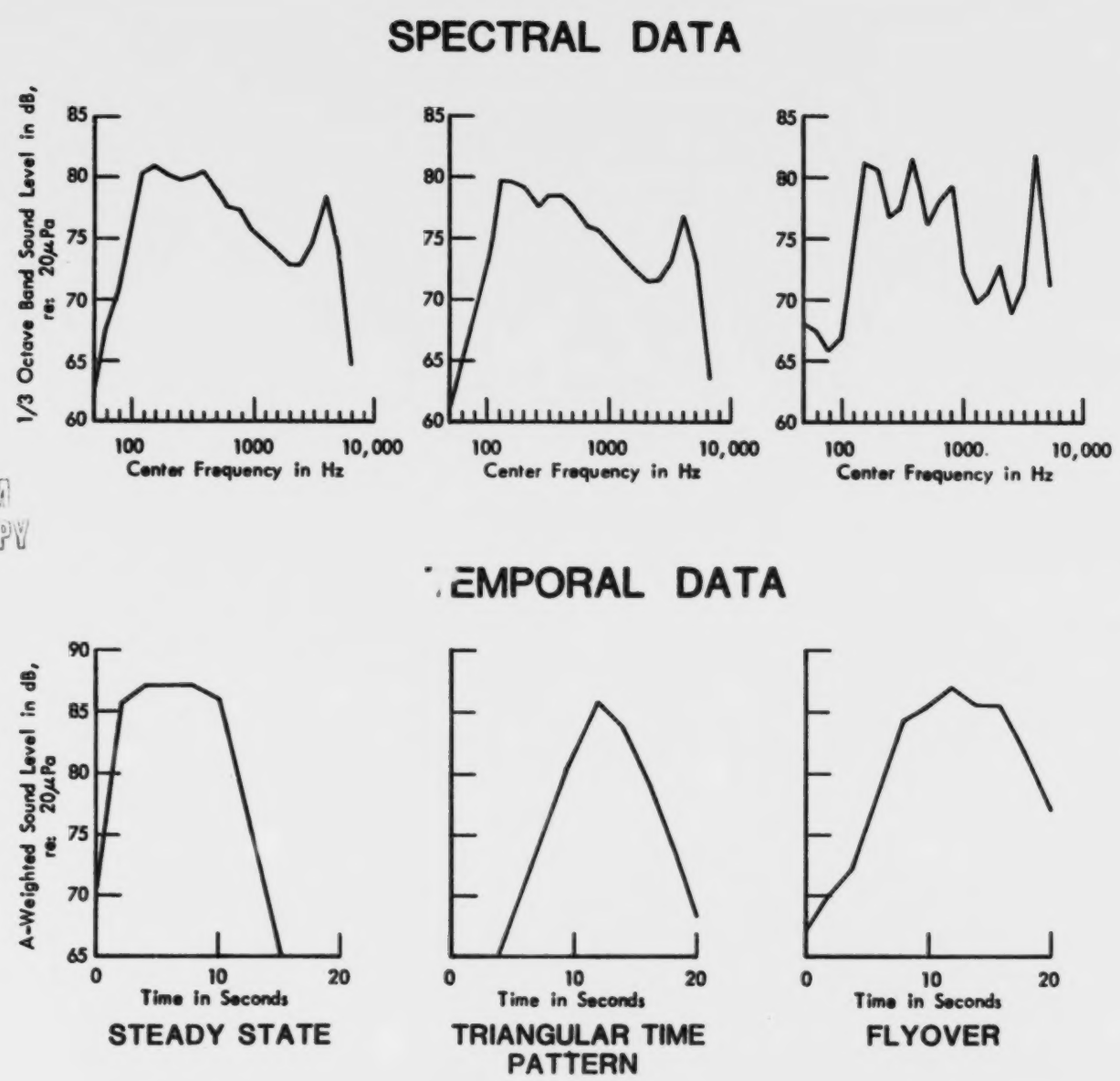
were performed on the data. Visual observations are presented first, with statistical manipulations motivated by these observations presented later.

Figure 8 shows the spectral and temporal composition of the prerecorded background noise intrusions. The three signals are presented here at their highest playback level. Recall that each signal was also presented at levels 10 and 20 dB lower. The signals were chosen to be relatively broadband in nature without significant spectral or temporal irregularities. The prominence in the spectrum at 4 kHz is an artifact of the headset response and because of its relatively high frequency is unlikely to have any substantial effects on the speech level.

Figure 9 shows the entire 286 second time history of the background noise and the measured speech level (V3 sound level) of a typical subject. Note that the background noise record is discontinuous showing levels for only the individual noise intrusions and not the dead time in between. The sound heard by the subjects during these intervals amounted to only a low level tape hiss (less than 35 dB[A]).

Generally, the relationship between speech and background levels appears quite good. As background levels rise and fall one can observe a complementary change in speech level as well, especially when the background exceeds 65 dB(A). At lower background levels it is more difficult to observe

MICROFILMED FROM
BEST AVAILABLE COPY



MICROFILMED FROM
BEST AVAILABLE COPY

FIGURE 8. SPECTRAL AND TEMPORAL DATA FOR BACKGROUND NOISE SIGNALS USED IN READING TEST

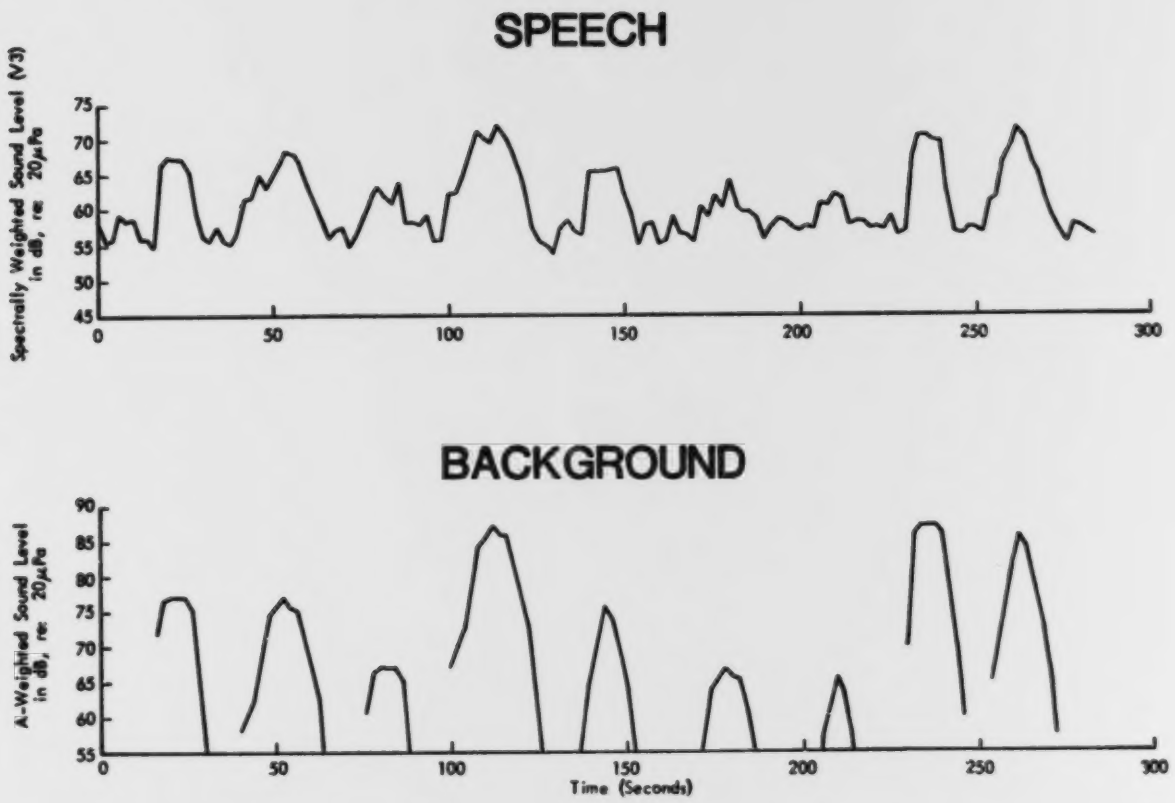


FIGURE 9. BACKGROUND NOISE AND SPEECH LEVELS OF INDIVIDUAL SUBJECT WHILE READING

a cause/effect relationship since the speech levels are approaching lower limiting normal conversational levels. The results shown in this figure suggest that good correlations between speech and background are likely to exist.

Straightforward correlations may be obtained by observing the speech and background levels at the points in time when each background noise intrusion reaches its maximum level. This provides 18 data points upon which trends may be established for each subject. Figure 10 shows this relationship for each subject by showing the speech level (V3 sound level) as a function of the background noise level (A-level). Least squares regression lines have been fitted to the data and slopes range from 0.28 to 0.49 with correlation coefficients (r) ranging from 0.65 to 0.96. Note that the steady state background intrusions (open circles) do not evoke systematically different speech levels from those of the other signals.

Figure 11 shows a more detailed view of one subject's data. This figure, similar to Figure 10, presents the relationship between speech and background noise levels at 2 second intervals during the six time-varying noise intrusions (data from the intervals between intrusions is omitted since it may be outside the range of the linear relationship). The slope of the dashed least squares fit line is not substantially different from that of the same subject in Figure 10, suggesting that only a limited number of data points may be necessary to establish such relationships.

MICROFILMED FROM
BEST AVAILABLE COPY

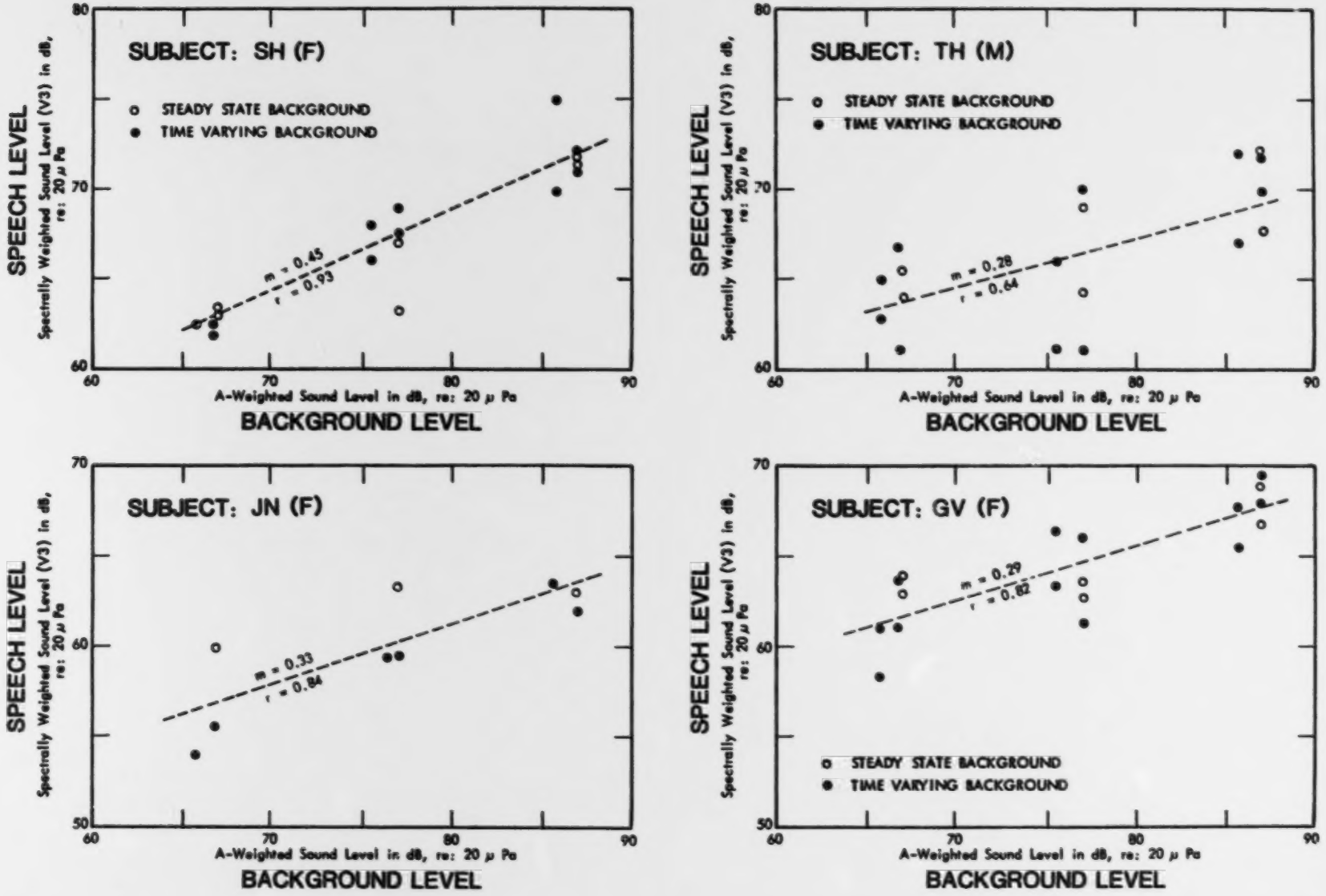
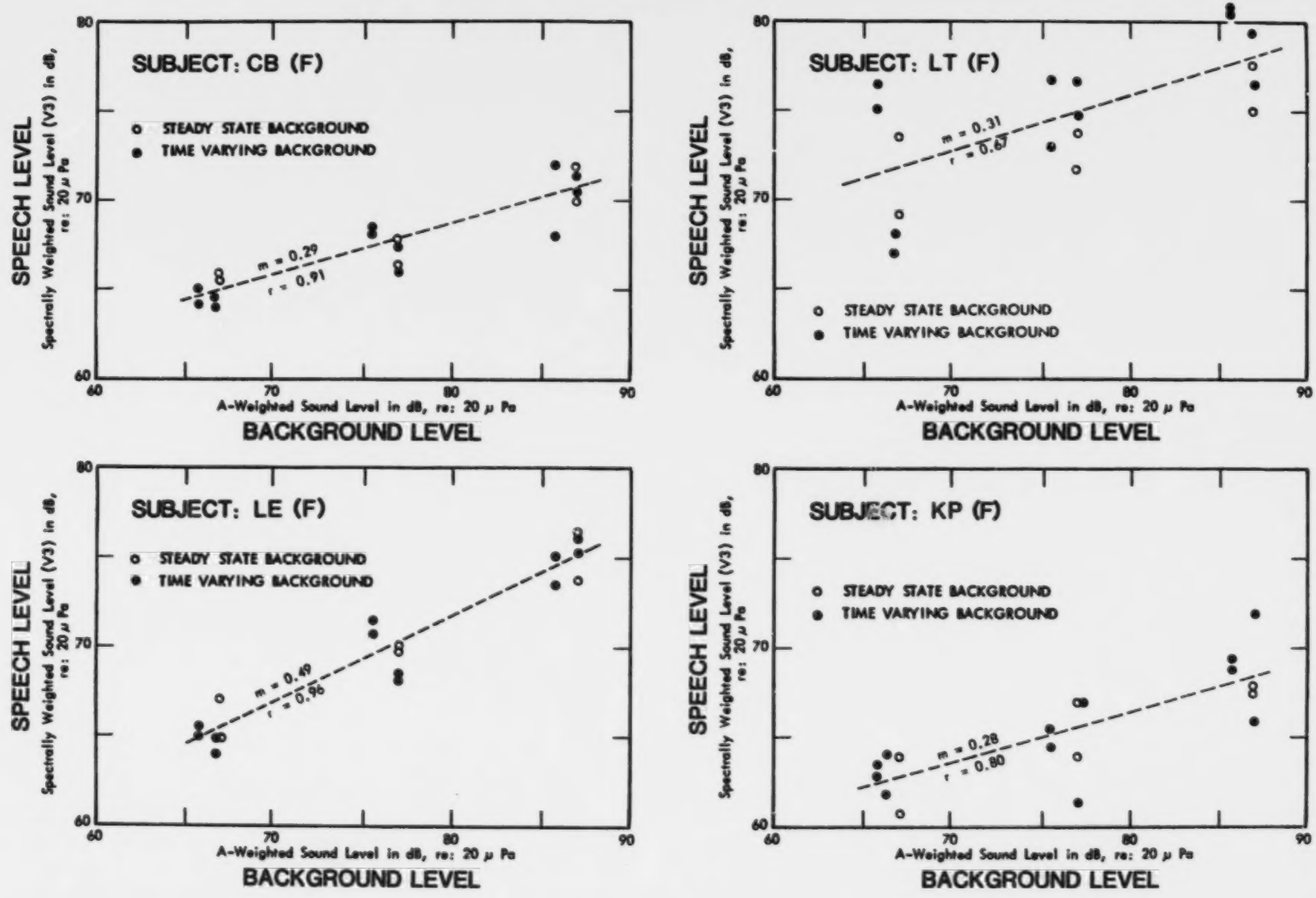


FIGURE 10. OBSERVED RELATIONSHIP BETWEEN SPEECH LEVELS AND BACKGROUND NOISE FOR INDIVIDUAL SUBJECTS

BEST AVAILABLE COPY
MICROFILMED



MICROFILMED FROM
BEST AVAILABLE COPY

FIGURE 10. (CONTINUED)

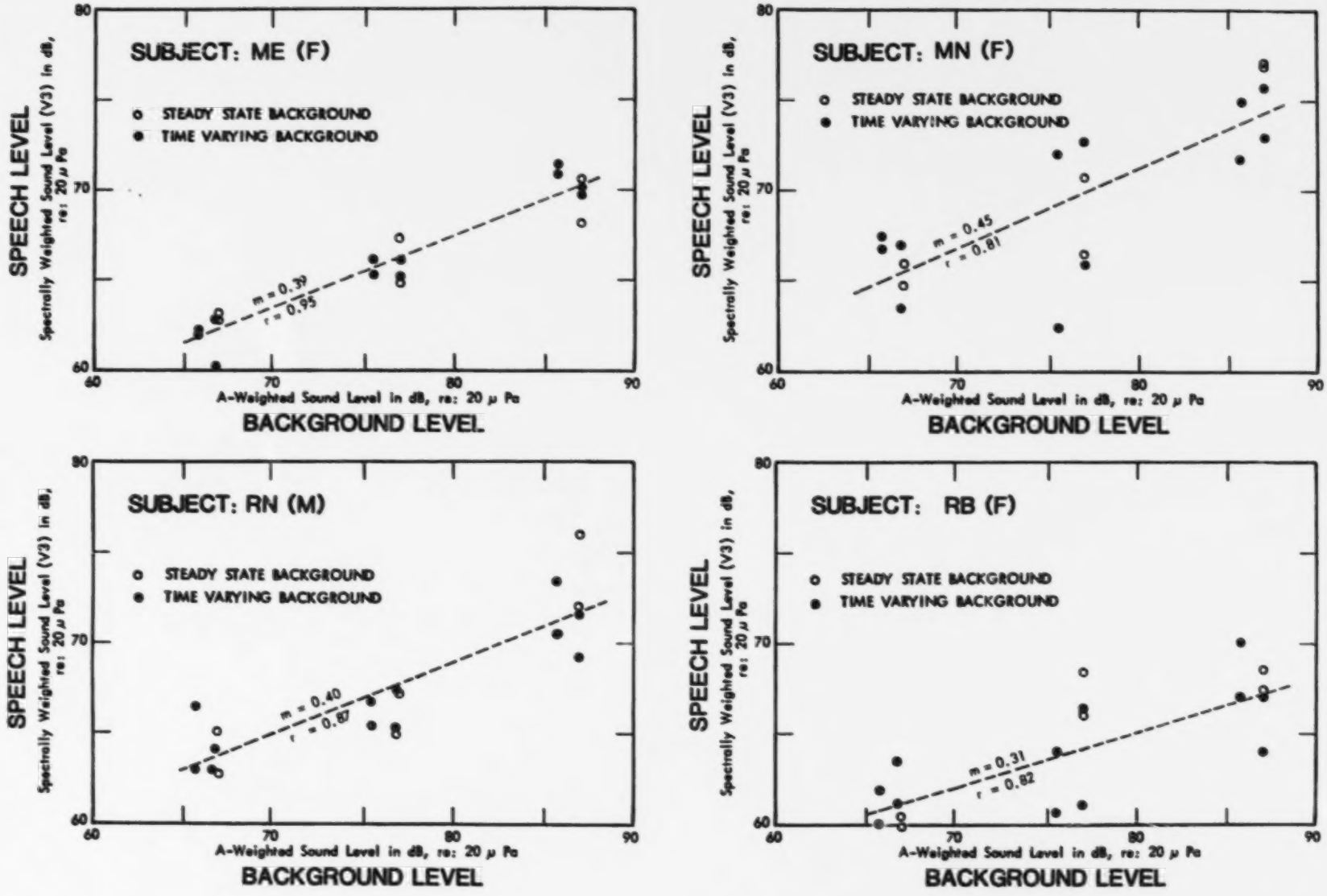
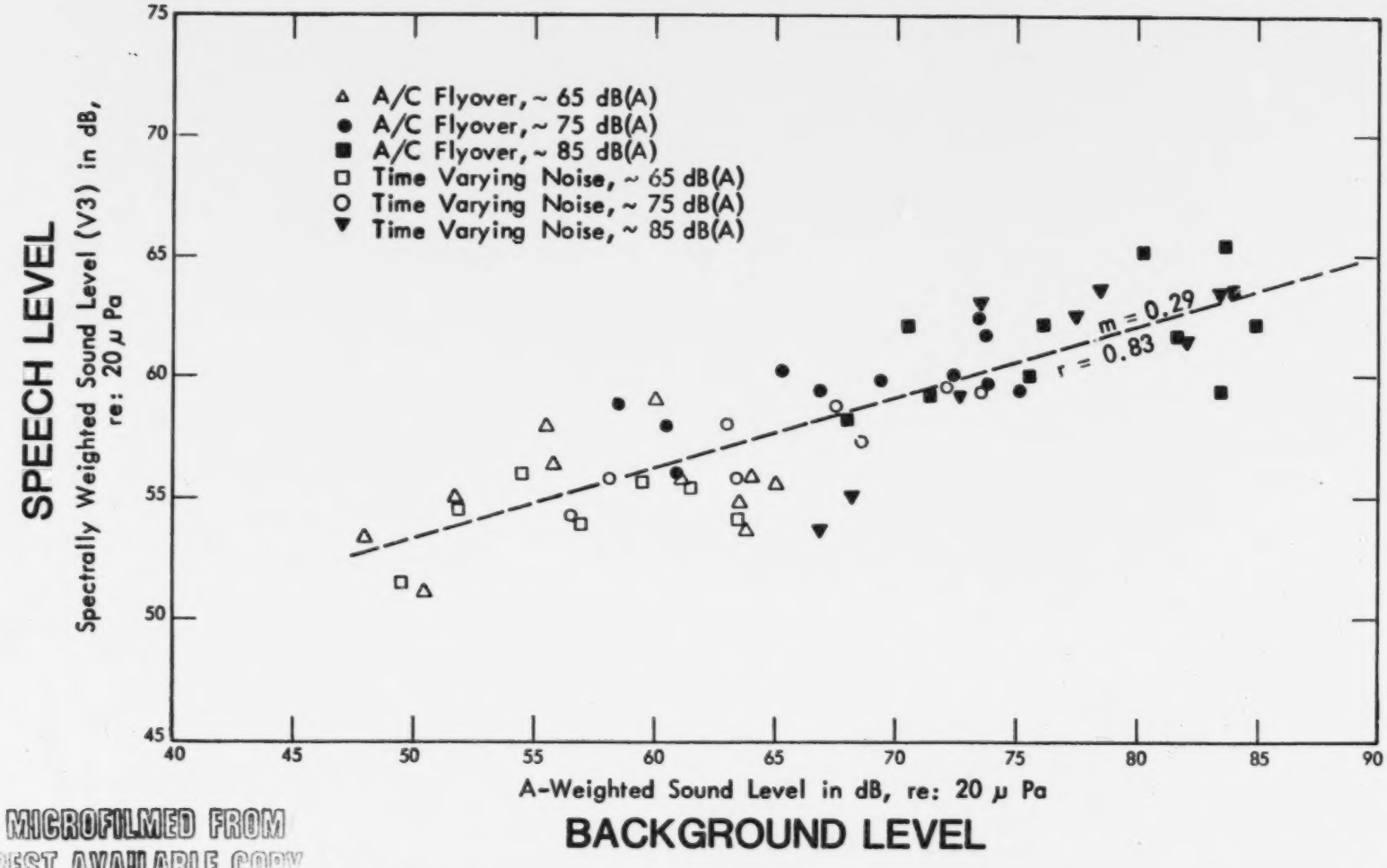


FIGURE 10. (CONTINUED)

MICROFILMED FROM
BEST AVAILABLE COPY



MICROFILMED FROM
BEST AVAILABLE COPY

FIGURE 11. OBSERVED RELATIONSHIP BETWEEN SPEECH AND BACKGROUND NOISE LEVELS AT 2 SECOND INTERVALS DURING NOISE INTRUSIONS (SUBJECT JN)

An admittedly second order effect (but nonetheless worth exploring) is the extent to which a temporal shift between the speech and background noise might affect the correlation between the two variables. For example, do people anticipate changes in the background noise and adjust their vocal effort ahead of time? Or do they have difficulty anticipating and in fact speak at levels commensurate with background levels a few moments earlier. To shed light on this question the background noise and one subject's speech data were reanalyzed using a one second (instead of two second) averaging time to obtain a greater time resolution. A cross-correlation analysis between the speech and background was then performed, with time displacements (τ) ranging from -9 to +9 seconds in one second intervals. Figure 12 shows the results of the analysis. The vertical axis shows the correlation coefficient (r), while the horizontal axis shows τ, the amount of time by which the speech leads or lags the background. The results of the analysis agree with commonsense expectations ..... the speech <u>lags</u> the background, but not by a large amount (about 0.5 seconds). This observation suggests that complex time delays are unnecessary for speech/background comparisons since substantial improvements in correlation are unlikely to occur.

Perhaps one of the most promising findings of this study is the ability of a relatively simple throat mounted vibration transducer to predict the speech sound level obtained from a conventional air microphone. Figure 13 shows the observed relationship between the air microphone

MICROFILMED FROM
BEST AVAILABLE COPY

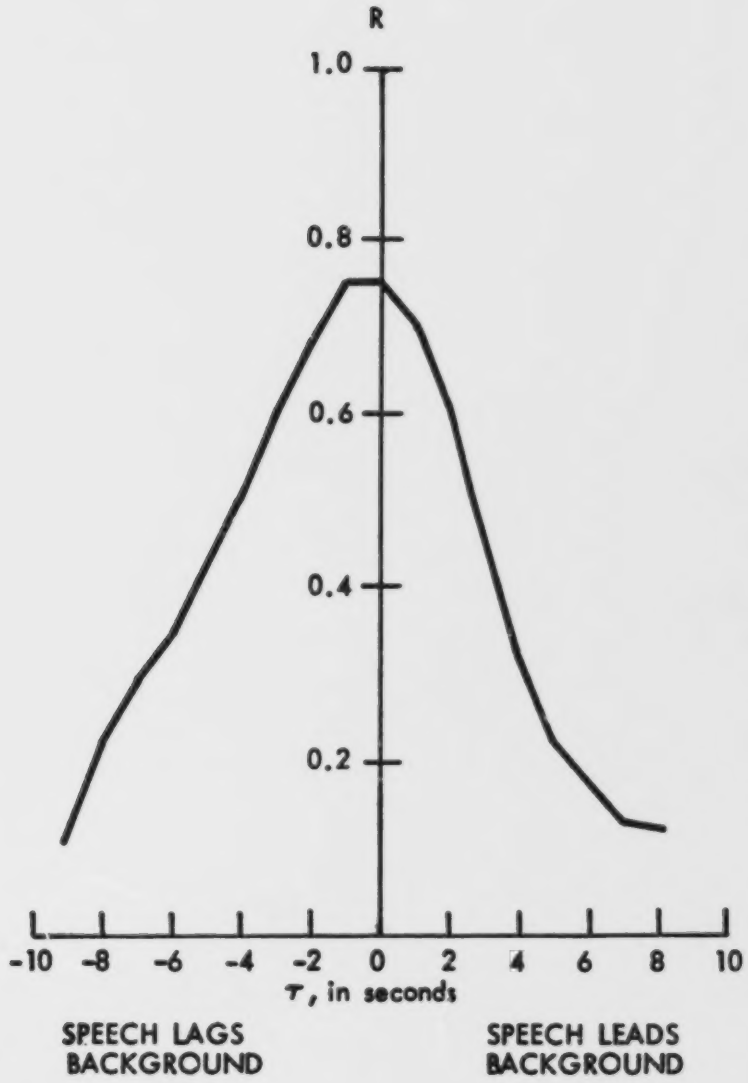


FIGURE 12. CROSS-CORRELATION OF BACKGROUND NOISE (>60 dB(A)) AND SPEECH LEVEL, V3

MICROFILMED FROM
BEST AVAILABLE COPY

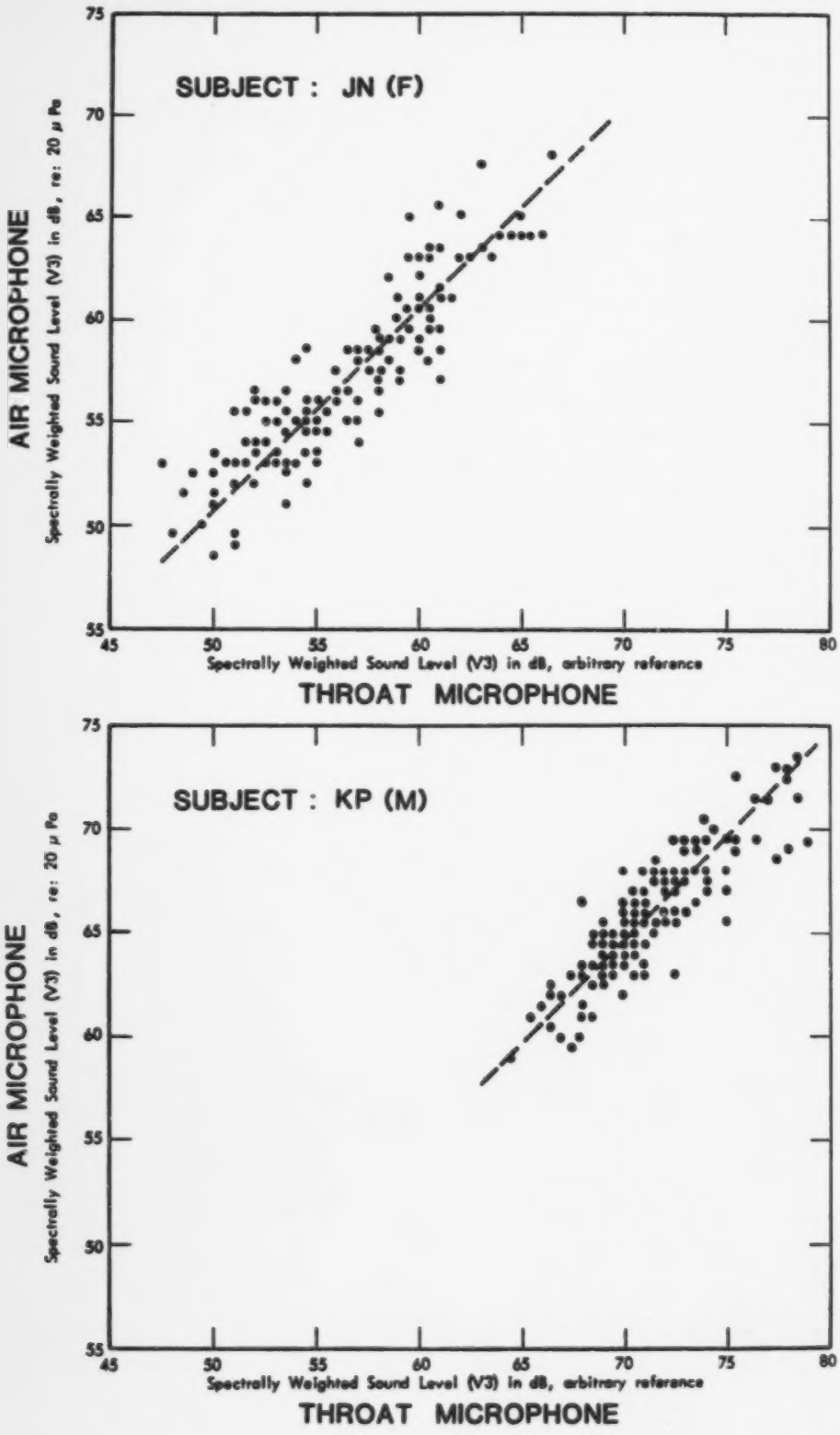

SUBJECT : JN (F)
AIR MICROPHONE
Spectrally Weighted Sound Level (V3) in dB, re: 20 μ Pa
Spectrally Weighted Sound Level (V3) in dB, arbitrary reference
THROAT MICROPHONE
SUBJECT : KP (M)
AIR MICROPHONE
Spectrally Weighted Sound Level (V3) in dB, re: 20 μ Pa
Spectrally Weighted Sound Level (V3) in dB, arbitrary reference
THROAT MICROPHONE

FIGURE 13. OBSERVED RELATIONSHIP BETWEEN CONVENTIONAL AIR MICROPHONE AND THROAT MICROPHONE

MICROFILMED FROM
BEST AVAILABLE COPY

levels and those obtained from the throat microphone for two test subjects. Subject JN, a female, was not a particularly loud talker at normal voice levels, but raised her voice considerably during some of the higher level noise intrusions. In addition the throat microphone was not tightened snugly around her neck (to determine the amount of latitude available in the attaching of the transducer). Note that there is no evidence of a non-linear relationship throughout the range of her data. The dashed line in the figure is a least squares fit with a forced slope of unity. Even with the loose transducer fit, the standard deviation about the regression line is only 1.8 dB.

In contrast, subject KP, a male, spoke at a generally higher level and had his throat microphone attached more securely than subject JN. The closer coupling to the neck may account for the reduced variability about the regression line (standard deviation about the regression line is 1.5 dB). Equally important is the absence of any non-linear trends at the higher levels.

To put these standard deviations in perspective consider the nominal 1.9 dB standard deviations about the regression lines (Figure 3 thorugh 7) in predicting speech level from background noise level. The amount by which this σ might be inflated if a throat microphone had been used to quantify speech level instead of the air microphone may be estimated by a simple orthogonal vector sum (square root of the sum of the squares) of two σ's, the 1.9 dB from the air microphone speech vs. background relationship, and the worst case 1.8 dB from the throat microphone vs. air microphone relationship. The vector sum is 2.6 dB (only 0.7 dB

MICROFILMED FROM BEST AVAILABLE

greater than using an air microphone). Stated another way, only twice as many ($[2.6/1.9]^2$) data points would be needed with the throat microphone to predict speech levels with the same precision as a conventional air microphone.

In order to provide a general relationship between speech level and background noise all the observations in Figures 3 to 7 were combined and a least squares fit line computed. The solid line in Figure 14 shows the regression line. The dashed line shows the 95% confidence interval on the regression line (not the scatter of individual data points). Note that the slope is in good agreement with those of the individual subjects, suggesting that there are no unusual anomalies between the data of individual subjects.

VI. DISCUSSION

The results suggest that it is possible to make short-term (1 to 2 second) measurements of speech level for individual talkers. However because of the nature of speech, more than one measurement should be made in a given background noise in order to determine more accurately the actual speech level produced by an individual. For example, if a measurement is desired within a 95% confidence interval 1 dB wide, then 10 to 15 observations should be made.

Although it was originally anticipated that vowel equalization would be necessary, it does not appear that vowel equalization filters improve the stability of the speech measurement. This is probably because the relationship between the vowels changes as the vocal effort increases

MICROFILMED FROM
BEST AVAILABLE COPY

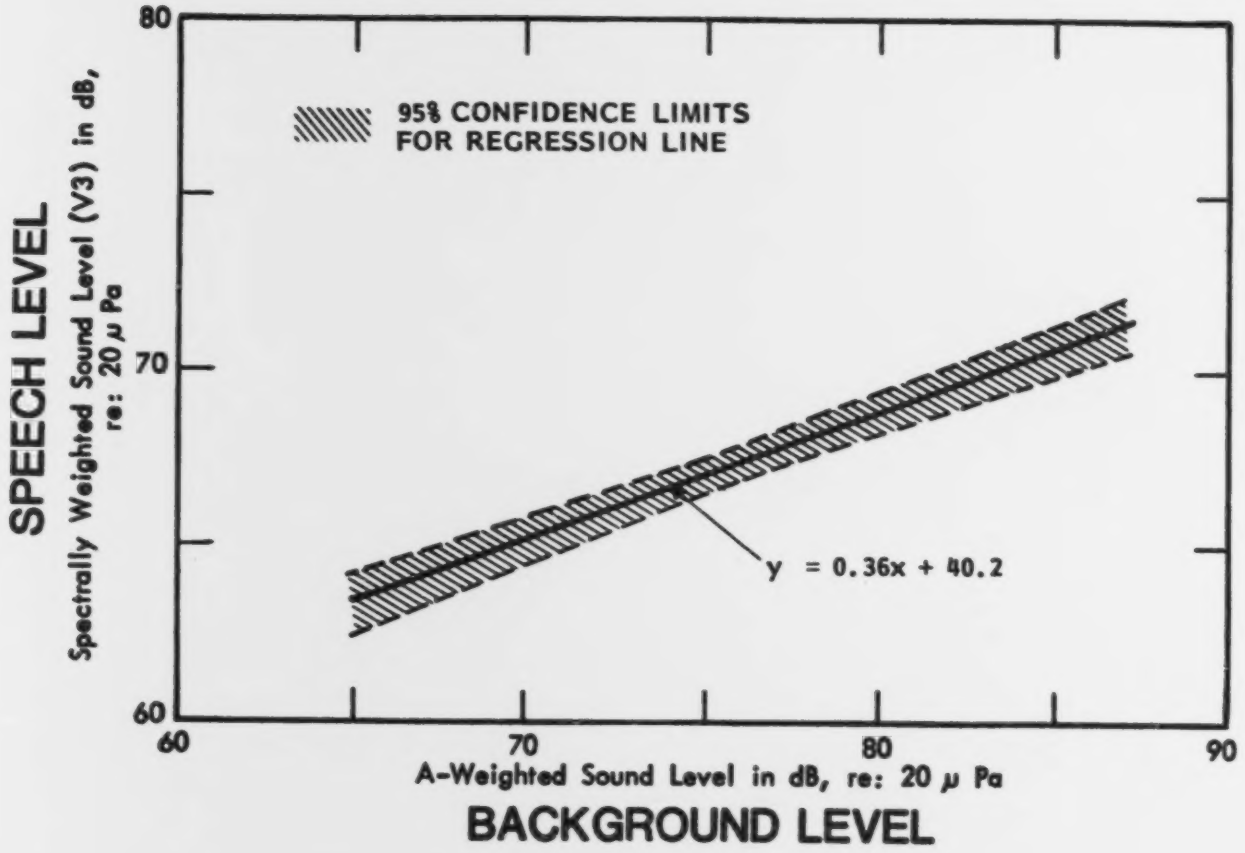


FIGURE 14. OBSERVED RELATIONSHIP BETWEEN SPEECH LEVELS AND BACKGROUND NOISE FOR ALL SUBJECTS

MICROFILMED FROM
BEST AVAILABLE COPY

as suggested in Figure 1. However, the bandwidth for measurement of speech is important in that the wider the bandwidth the more stable the measure. Further, the time constants necessary for short term measurement of speech are important especially if one is attempting to measure speech levels in the presence of a time varying noise such as aircraft flyover noise. If the time constants are too long, then the speech levels will be underestimated as the background noise increases and overestimated as the background noise decreases. On the other hand, if the time constants are too short, then the variability of the speech measure is too great to be of use.

It should be emphasized that other measures than overall or V3 weighted sound level may be used with equal accuracy. However, measures should be broadband in nature so that fluctuations in speech associated with certain vowels do not influence the overall measures of speech.

One word of caution should be made regarding long pauses in speech material. All of the speech material used in this study was either from readings or repetition of memorized phrases which contain no long pauses. However, in actual conversations sometimes long pauses do exist and care should be taken not to allow these pauses to influence the measure of speech. The equivalent peak level is probably the best current measure for coping with these long pauses and although the measure indicated greater variability than other measures, it would pro-

MICROFILMED FROM
BEST AVAILABLE COPY

bably be an improvement over the other measures if the speech material contained long pauses.

Having obtained a reasonable short term measure of speech, its reliability was further tested using talked in accordance with the protocol discussed in Phase II of this investigation. Both steady state and time-varying noises were employed, however, no difference in the speech levels produced was observed between data obtained with steady state or time-varying noise. The overall relationship of speech level and background noise based on slope of Figure 14 indicated that people automatically raise their voice 3-1/2 dB for each 10 dB of increase in background noise over the range of background noise levels from 65 to 85 dB. This relationship is comparable to that observed by other investigators (ref. 10, 11).

Some testing was done using a throat microphone (vibration pickup) which provided a good signal-to-noise ratio and also was highly correlated with the conventional acoustical measurements. This should provide a reasonable method for obtaining speech measurements in the presence of high background noise levels.

Using the speech measurements obtained with the technique mentioned in this report would allow estimations of intelligibility in the presence of various background noises and in particular on a moment-by-moment basis while such transient events as aircraft flyovers are occurring.

MICROFILMED FROM
BEST AVAILABLE COPY

However, all intelligibility estimates using articulation index have been based on normal voice level and some caution should be exercised since the intelligibility using a raised voice may not be as good as that using the estimation process since it is more difficult to enunciate when using a raised voice.

VII. CONCLUSIONS

The following conclusions may be drawn as a result of the analysis of speech measurements obtained and summarized under this investigation.

1) It is possible to estimate the long term RMS level of speech for continuous discourse using two second samples. These samples will be distributed with a standard deviation of 1.5 dB, provided no long pauses in speech material exists.

2) Vowel equalization techniques appear to provide no improvement over the overall sound pressure level measurements of speech.

3) For determining speech levels in noisy environments, use of a throat microphone provides results which correlate very well with conventional microphone measurements.

4) Preliminary data using the short-term speech measurement system indicates that people automatically raise their

MICROFILMED FROM
BEST AVAILABLE COPY

voice about 3-1/2 dB for each 10 dB increase in background level. Individuals differ in absolute level, but do not appreciably differ in the rate at which they increase their voice level with increasing background noise. This finding is in agreement with other speech level studies.

MICROFILMED FROM
BEST AVAILABLE

REFERENCES

1. Dunn, H. K., White, S. D., "Statistical Measurements on Conversational Speech", J. Acoust. Soc. Am. 11, 278-288, 1940.

2. Beranek, L. I., "The Design of Speech Communication Systems", Proc. I.R.E. 35, 880-890, 1947.

3. ANSI American National Standard Methods for the Calculation of the Articulation Index. ANSI S3.5 (1969).

4. Benson, R. W., Hirsh, I. J., "Some Variables in Audio Spectrometry", J. Acoust. Soc. Am. 25, 499-505, 1953.

5. French, N. R., Steinberg, J. C., "Factors Governing the Intelligibility of Speech Sounds", J. Acoust. Soc. Am. 19, 90-119, 1947.

6. Pearsons, K. S., Bennett, R. L., and Fidell, S., "Speech Levels in Various Noise Environments", EPA-600/1-77-025, 1977.

7. Brady, P. T., "Equivalent peak level: A threshold-independent speech-level measure", J. Acoust. Soc. Am. 44, 695, 1968.

8. Steeneken, H. J. and Houtgast, T., "Comparison of some methods for measuring speech levels", IZF 1978-22, 38, Institute for Perception TNO (The Netherlands).

9. Peterson, G. E., and Barney, H. L., "Control Methods Used in a Study of the Vowels", J. Acoust. Soc. Am. 24, 175-184, 1952.

10. Lane, H., Tranel, B., "The Lombard Sign and the Role of Hearing in Speech", J. Sp. & H. Res. 677-709, 1971.

11. Heusden, E. V., Plomp, R., and Pols, L. C., "Effect of Ambient Noise on the Vocal Output and the Preferred Listening Level of Conversational Speech", Applied Acoustics 12, 1979.

12. ANSI American National Standard Specifications for Audiometers. ANSI S3.6 (1969).

MICROFILMED FROM
BEST AVAILABLE COPY

Blank Page

MICROFILMED FROM
BEST AVAILABLE

APPENDIX A

EQUIVALENT PEAK LEVEL (EPL)

MICROFILMED FROM
BEST AVAILABLE COPY

APPENDIX A - EQUIVALENT PEAK LEVEL (EPL)

The equivalent peak level, or EPL, is a speech level measure which is based on the empirical finding that the logarithm of the instantaneous absolute magnitude of speech samples is (nearly) uniformly distributed between any arbitrary threshold value T, and a peak, p. The EPL is an estimate of p. The estimate is derived by choosing a threshold T that is high enough to clear noise, and low enough to fall below most of the speech samples; then, one measures the average of the square of only those voltage samples that clear the threshold. This produces an "average RMS". If one knows the measured RMS and the chosen threshold one can deduce the value of p, called EPL. For true log-uniform distributions the same value of p will result no matter what threshold was chosen, i.e. EPL is threshold independent. In the operational definition of EPL, an empirical correction is applied to compensate for departures in speech waveforms from log-uniformity in the higher ranges of speech level.

The operational definition of EPL is as follows:

1. Choose a threshold T between the noise level and the peaks of the speech. (A thresholf 15 to 20 dB below the expected EPL works well; but the choice is not critical, since threshold independence of EPL typically holds over a threshold range of over 30 dB).

2. Measure the average volts squared for only those voltages that exceed T.

MICROFILMED FROM
BEST AVAILABLE COPY

3. Convert this measurement to a decibel measure such as dBm, dBV, dB20uPa, etc.

4. Obtain D = rms minus T. (T must also be expressed in decibels). Then compute a value Δ from the function shown below (illustrated in graphical form in Figure A-1).

   a. if D < 6.75, set Δ = (D-2.75)/0.4[(a)]
   b. if 6.75 < D < 13.5; set Δ = D/0.675
   c. if D > 13.5, set Δ = (D+2.88)/0.819

5. EPL = T+Δ

The EPL algorithm was implemented on a Digital Equipment Corporation PDP-8 computer equipped with a 12-bit analog-to-digital (A/D) converter. A nominal A/D sampling rate of 1 kHz was used to acquire instantaneous speech samples. Intentional jitter was introduced into the sampling rate to minimize possible discrete frequency biases.

The sample interval for which the computer would calculate and report an EPL value was under operator control. For this study EPL's were reported at 1-second intervals in Phase I and at 2-second intervals in Phase II.

Seven fixed thresholds were used, each 6 dB apart, with the highest threshold 6 dB down from the maximum range of the A/D converter. Playback gain was adjusted so that the speech waveform peaks occurred in the top 6 dB

(a) Step 4a of the current operational definition was revised in 1969, and differs slightly from that published by Brady in 1968 (ref. 7).

MICROFILMED FROM
BEST AVAILABLE COPY

of the converter range. Starting with the highest threshold, the EPL was computed. If certain computational criteria were not met the next lowest threshold was used and a new EPL value computed. The following algorithm was employed to select the proper EPL to be reported.

1. Compute the EPL for the currently selected threshold. If there are sufficient samples above "T" (i.e., more than 50% of the total sample size for the observation interval) and if the EPL is at least 12 dB above T then print this EPL value. Otherwise, select the next lower threshold (i.e. reduce it by 6 dB) and repeat this step.

2. If upon reaching the lowest threshold the criteria are still not met, then an under range condition is reported in place of the EPL.

At the conclusion of the entire speech recording the computer printed a time history of the computed EPL values.

MICROFILMED FROM
BEST AVAILABLE COPY

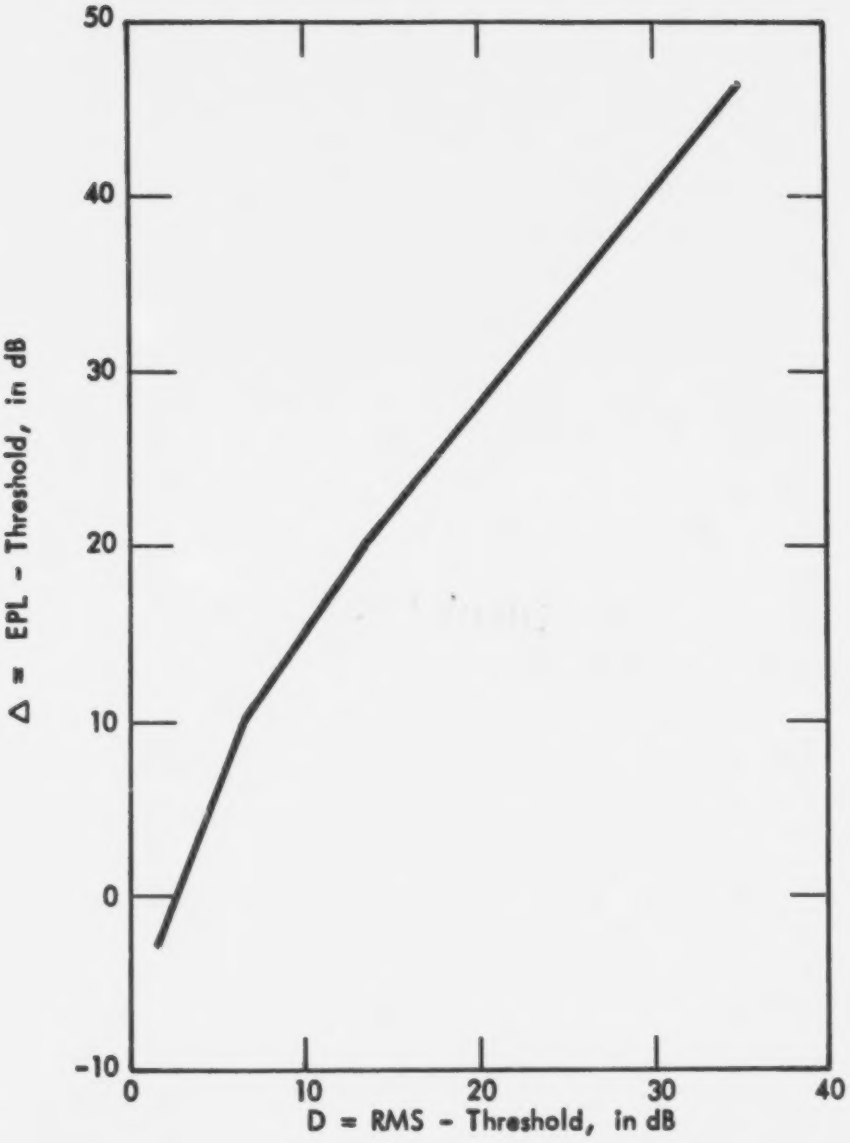


FIGURE A-1. RELATIONSHIP BETWEEN PEAK AND RMS ABOVE THRESHOLD

MICROFILMED FROM
BEST AVAILABLE COPY

Blank Page

APPENDIX B

INSTRUMENTATION

MICROFILMED FROM
BEST AVAILABLE COPY

APPENDIX B - INSTRUMENTATION

A. Speech Recording Instrumentation (Pearsons, ref. 6)

Speech samples acquired by Pearsons (ref. 6) were recorded on magnetic tape for subsequent playback and analysis. A block diagram of the recording system is shown in Figure B-1. The microphone and preamplifier were located inside an anechoic chamber at normal incidence to (and 1 meter from) the speaker and connected by approximately 50 feet of cable to a sound level meter and tape recorder in an adjacent laboratory. The microphone was a Bruel & Kjaer Type 4133 1/2-inch condenser microphone with a General Radio Type 1560-P42 microphone preamplifier. A Bruel & Kjaer Type 2203 precision sound level meter was used as a decading amplifier in front of an Ampex AG-350 2-channel, 1/4-inch magnetic tape recorder. A Bruel & Kjaer Type 4220 pistonphone calibrator was used for absolute level calibration.

B. Speech Recording and Background Noise Playback Instrumentation (Current Study)

Test participants recited prepared text in an anechoic environment while listening to periodic noise intrusions over earphones with open cellular foam cushions. The speech was recorded on magnetic tape from (1) a microphone at normal incidence to (and 1 meter from) the speaker and (2) a throat mounted vibration transducer. Figure B-2 shows a block diagram of the instrumentation used.

MICROFILMED FROM
BEST AVAILABLE COPY

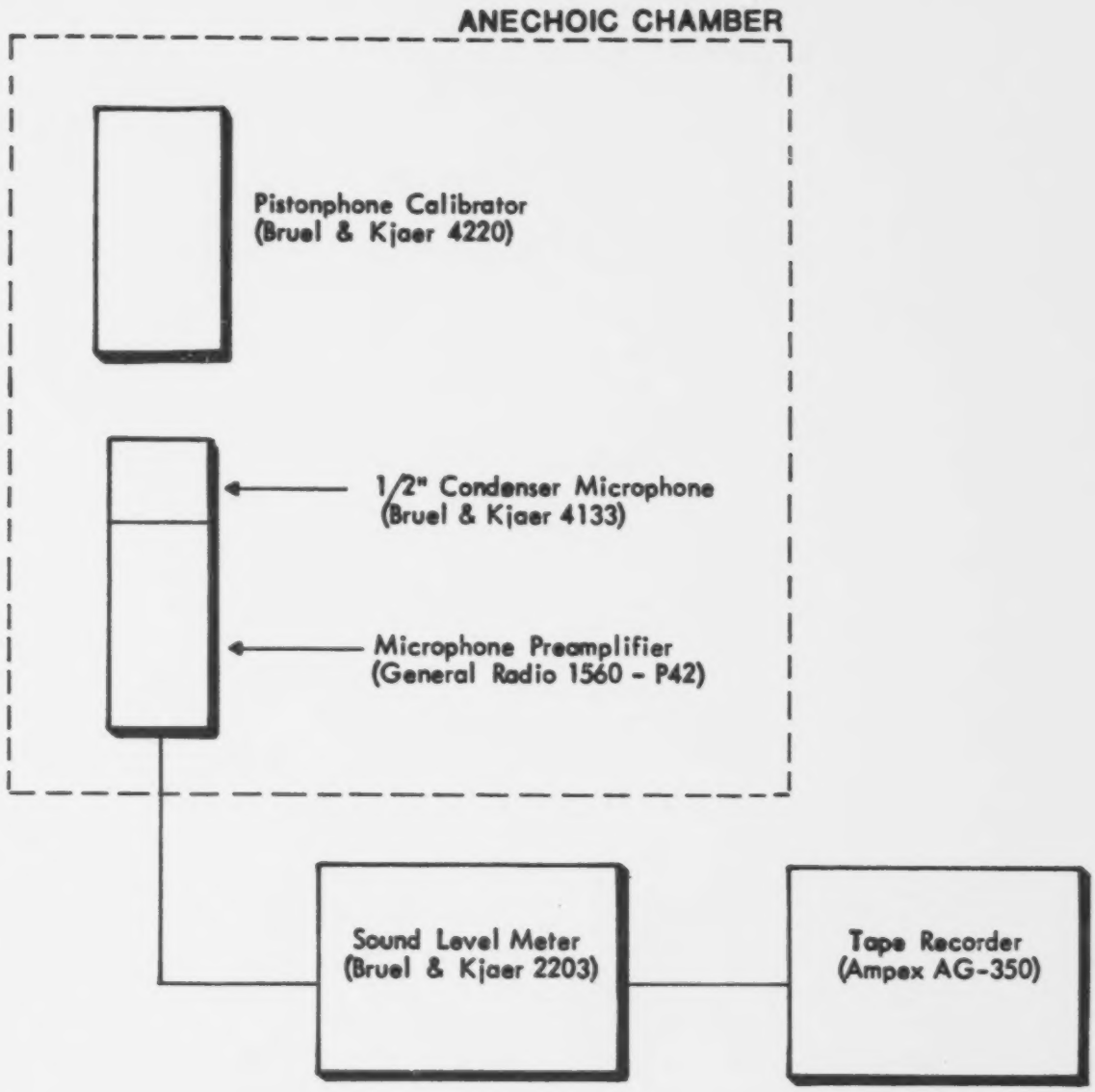


FIGURE B-1. BLOCK DIAGRAM OF VOICE MEASUREMENT EQUIPMENT

MICROFILMED FROM
BEST AVAILABLE COPY

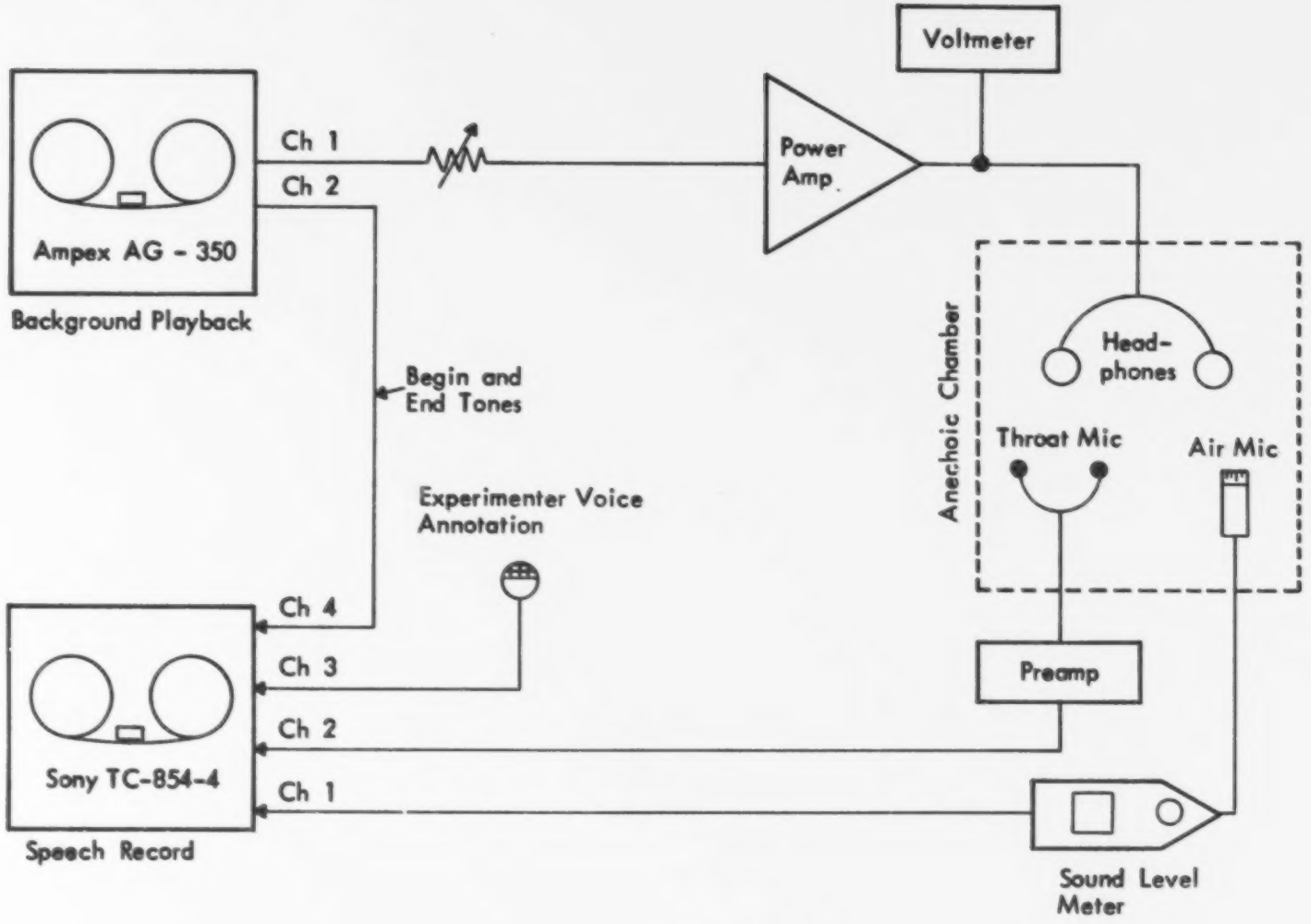


FIGURE B-2. BLOCK DIAGRAM OF BACKGROUND NOISE GENERATION AND SPEECH RECORDING INSTRUMENTATION

MICROFILMED FROM
BEST AVAILABLE

A prerecorded monaural background noise tape of nominal 5-minute duration was played back on an Ampex AG-350 1/4-inch, 2 channel magnetic tape recorder through a Daven 0.1 dB step attenuator, a Pioneer Model SX-450 Stereo Receiver and Sony Model MDR-7 stereo headphones. The stereo receiver divided the single channel input from the tape recorder equally between the left and right earphones. A Balentine Model 320A True RMS Voltmeter was connected to the earphone lines for calibration purposes. Calibration involved setting the voltage level of a prerecorded 1000 Hz tone (at the beginning of the background noise tape) and ensuring that the two earphone channels were in balance.

The second channel of the background noise tape contained prerecorded cue tones (500 millisecond burst at 2 kHz) at the beginning and end of the nominal 5 minute background noise segment. These tones were used in subsequent data analysis to synchronize the background noise and speech recordings.

The recording microphone was a Bruel & Kjaer Type 4133 1/2-inch condenser microphone with a Bruel & Kjaer Type 2615 microphone preamplifier (and a Bruel & Kjaer Type 2801 power supply). The output of the preamplifier was connected to a Bruel & Kjaer Type 2203 precision sound level meter which acted as a decading amplifier in front of a Sony Model TC-854-4 4-channel, 1/4-inch magnetic tape recorder. Microphone calibration was

MICROFILMED FROM
BEST AVAILABLE COPY

by a Bruel & Kjaer Model 4220 Pistonphone, with the pistonphone signal being recorded directly on the magnetic tape.

The throat transducer was powered and preamplified by inhouse built hardware. The preamplifier output was recorded on channel 2 of the Sony recorder. No attempt was made at absolute calibration of the transducer.

An annotation microphone was connected to channel 3 of the recorder for the experimenter to document pertinent data regarding each recording. Channel 4 recorded the beginning and end synchronization cue tones from channel 2 the background noise tape so that speech sound levels and background noise could be synchronized and compared at any instant in time.

C. Spectral Analysis Instrumentation

Analysis of various frequency weighted sound level measures was performed by a mini computer and one-third octave band real-time spectrum analyzer. Figure B-3 shows a block diagram of the instrumentation.

Signals on magnetic tape were played back from the tape transport on which they were recorded (Sony TC-854-4 or Ampex AG-350) into a Hewlett-Packard Model 8054A One-Third Octave Band Real Time Audio Spectrum Analyzer. The analyzer was set to "RMS SLOW" sound level meter characteristics and meets IEC 179 precision sound level meter specifications (measured RC time constant in all

MICROFILMED FROM
BEST AVAILABLE COPY

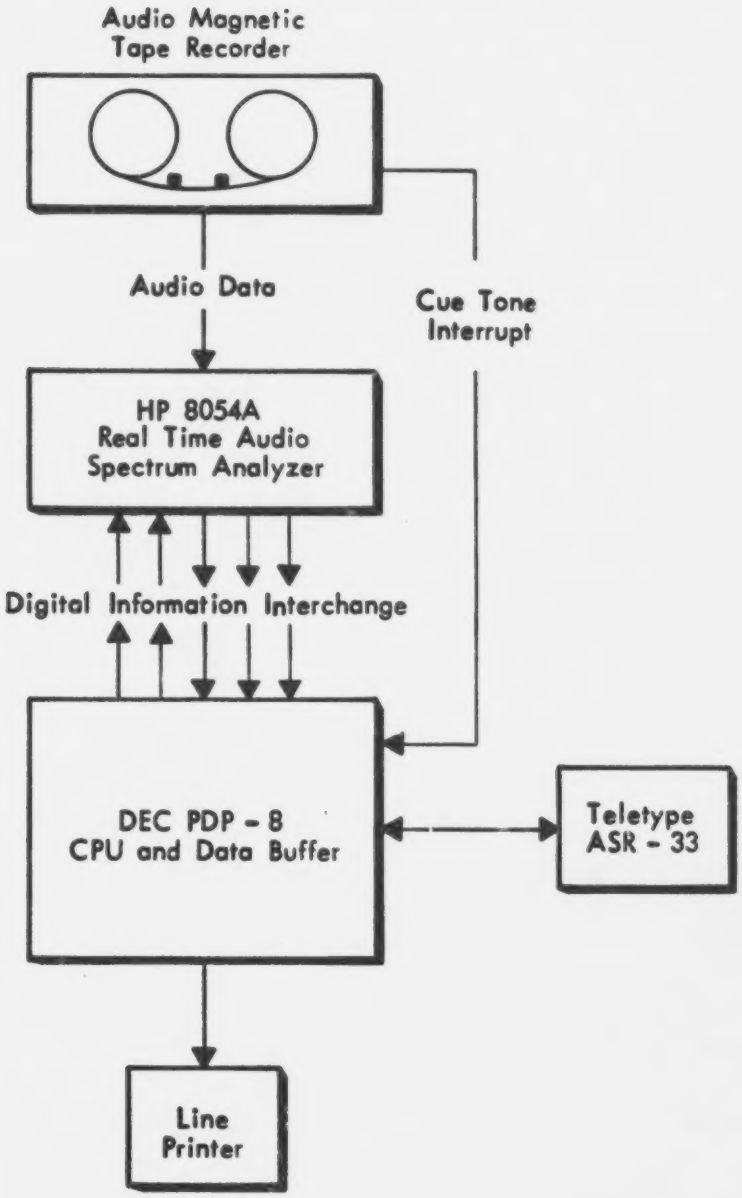


FIGURE B-3. REAL TIME SPECTRAL ANALYSIS SYSTEM

MICROFILMED FROM
BEST AVAILABLE COPY

frequency bands is 1100 msec ± 10%). Band center frequencies range from 50 to 10,000 Hz. Crest factor capability is 20 dB with less than 1 dB error.

The spectrum analyzer is interfaced to a Digital Equipment Corporation (DEC) PDP-8 digital computer. The prerecorded cue tones delimit the beginning and end of the tape segment to be analyzed. The spectrum analyzer performs continuous RC signal averaging independent of computer control. Upon request from the computer the analyzer digitizes the RMS sound level in a specified band and transmits this data. Analysis commences with the first cue tone. The computer requests a digitized one-third octave band spectrum from the analyzer at regular periodic intervals (one per second in this study) and stores them in memory until the second delimiting tone is encountered.

The computer calculates the various frequency weighted measures (e.g., A-level) by appropriately weighting and summing the one-third octave band levels and reports the time history of these measures on hard copy output.

D. Equivalent Peak Level (EPL) Analysis Instrumentation

Equivalent peak level (EPL) analyses were performed by a mini computer and analog-to-digital (A/D) converter. A block diagram of the instrumentation is shown in Figure B-4.

Signals on magnetic tape were played back from the tape transport on which they were originally recorded

MICROFILMED FROM
BEST AVAILABLE COPY

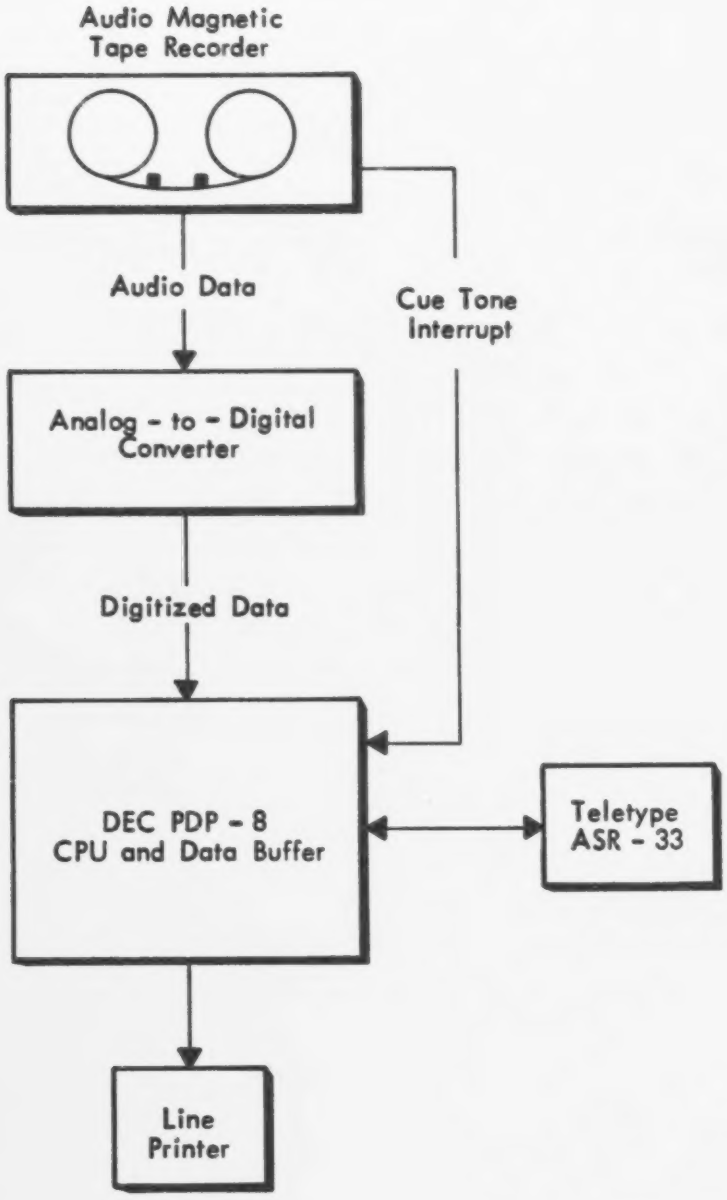


FIGURE B-4. REAL TIME EQUIVALENT PEAK LEVEL (EPL) ANALYSIS SYSTEM

MICROFILMED FROM
BEST AVAILABLE COPY

into an analog-to-digital converter and Digital Equipment Corporation PDP-8 digital computer.

A pair of cue tones on the magnetic tape delimited the data segment to be analyzed. Their presence is sensed by external hardware which signals the computer. The audio signal waveform was digitzed at a nominal 1 kHz rate (intentional intersample jitter was introduced to avoid discrete frequency biases) and the EPL calculated at 2-second intervals as per Brady (ref. 7). A time history of the computed EPL values is reported on hard copy output. Communication between the operator and computer is provided by a Teletype Model ASR-33.

MICROFILMED FROM
BEST AVAILABLE COPY

APPENDIX C

INSTRUCTIONS TO TEST SUBJECTS

MICROFILMED FROM
BEST AVAILABLE COPY

APPENDIX C - INSTRUCTIONS TO TEST SUBJECTS

Test subjects were given written instructions regarding their participation at the beginning of the speech recording session. Figure C-1 shows these instructions.

MICROFILMED FROM
BEST AVAILABLE COPY

The task you are about to undertake will involve reading the attached newspaper article aloud, as though you wanted to communicate with a person a few feet away at the dummy's location. Your voice will be recorded while you are reading.

While you are reading you will be wearing earphones through which you will occasionally hear various sounds. Imagine that these sounds are occurring in your living room, and you want the person to whom you are speaking to understand everything you say.

Thank you for your help.

FIGURE C-1. INSTRUCTIONS TO TEST SUBJECTS

MICROFILMED FROM
BEST AVAILABLE COPY

Blank Page

APPENDIX D

PREPARED TEXT

MICROFILMED FROM
BEST AVAILABLE COPY

APPENDIX D - PREPARED TEXT

Test subjects read two sets of prepared text during the speech recording session. Excerpts from newspaper articles, the text was retyped in large print to facilitate reading and is shown in Figures D-1 and D-2.

MICROFILMED FROM
BEST AVAILABLE COPY

FIVE COTTONWOOD TREES FOR FOUR LITTLE GIRLS AND THEIR MOTHER WERE PLANTED IN SOUTH DAKOTA BY A PROUD FATHER ONE HUNDRED YEARS AGO. THE LAKE THEY BORDERED HAS DRIED UP, BUT BECAUSE ONE OF THE GIRLS REMEMBERED AND WROTE OF THAT AND OTHER EVENTS IN HER PIONEER CHILDHOOD, THE "WEST" OF HOMESTEADERS IS STILL ALIVE AND READY FOR COMPANY.

THE TV VERSION OF "LITTLE HOUSE ON THE PRAIRIE" MAY HAVE BEEN PARTLY RESPONSIBLE FOR THE CONTINUING INTEREST IN LAURA AND HER FAMILY, BUT IT HAS ALSO RAISED THE HACKLES OF BOOK FANS.

PREDICTABLY, SCRIPT CHANGES HAVE BEEN AT VARIANCE WITH THE FACTS; PEANUT BUTTER SHOWN BEFORE IT WAS INVENTED; THE NONEXISTENT TOWN OF WINOKA INSTEAD OF THE REAL DE SMET. SUCH ERRORS ARE CAMPFIRE TALK IN THE SUMMERTIME. THOSE WHO FLOCK TO THE SITES OF THE BOOKS KNOW BETTER.

FOR THE LAST 12 YEARS VISITORS HAVE WALKED THROUGH THE TINY ROOMS OF THE RESTORED SURVEYOR'S HOUSE WHERE THE WILDERS WINTERED A CENTURY AGO. DE SMET, SOUTH DAKOTA, POPULATION 1,500, HAS ONE HOTEL, TWO MOTELS AND A MAIN STREET, STILL AND FOREVER CALLED MAIN STREET.

IN SUMMER IT ALSO HAS ACRES OF CAMPSITES AND MAKES USE OF THE FACILITIES OF NEIGHBORING TOWNS AS WELL. THE OCCASION IS THE ANNUAL OUTDOOR PAGENT BASED ON LAURA'S FIFTH BOOK, "THE LONG WINTER", STARRING 25 RESIDENTS DRESSED IN THE COSTUMES OF THE LAST CENTURY.

THE 1981 DATES ARE JUNE 27 AND 28, JULY 4-5 AND JULY 11-12, AND THE PLACE IS THE ORIGINAL SITE OF THE INGALLS' HOMESTEAD, ABOUT A MILE SOUTHEAST OF THE PRESENT TOWN.

ALTHOUGH THE PAGENT CONTAINS SOME SCENES OF LAURA'S CHILDHOOD, THE CONCENTRATION IS ON THE HARSH WINTER OF 1880. ADULTS PAY $2, CHILDREN UNDER 12, $1 (PRESCHOOLERS FREE), AND ALTHOUGH THERE IS SOME PLANK SEATING, THE AUDIENCE IS ADVISED TO TAKE LAWN CHAIRS AND BLANKETS. THE HOUR-AND-A-HALF PERFORMANCE BEGINS AT 9 P.M. A MILE EAST OF DE SMET.

THE LAURA INGALLS WILDER MEMORIAL SOCIETY PROVIDES BOTH GUIDED AND SELF-DIRECTED TOURS TO 18 SITES MENTIONED IN THE BOOKS AND STAFFS THE SURVEYOR'S HOUSE AND THE LAST HOME OF THE INGALLS FAMILY (BUILT BY PA AND OCCUPIED FROM 1887 TO 1928, AFTER THEY MOVED OFF THE HOMESTEAD AND INTO TOWN).

THE HOUSE CONTAINS SUCH MEMORABILIA AS MA'S KEROSENE LAMP, CARRIE'S MUFF AND FUR COAT, PA'S TRUNK. A SECOND-

MICROFILMED FROM BEST AVAILABLE COPY

FIGURE D-1. "LITTLE HOUSE" TEXT READ BY TEST SUBJECTS

STORY BEDROOM IS FURNISHED WITH ARTICLES FROM THE CONNECTICUT HOME OF ROSE WILDER LANE, LAURA'S ONLY DAUGHTER. THE SURVEYOR'S SHANTY HAS BEEN RESTORED TO THE TIME OF THE INGALLS' STAY. THERE IS EVEN A COPY OF THE WHATNOT SHELF PA BUILT FOR MA SO LONG AGO.

THE SITE MAY BE VISITED FROM MAY FIRST TO SEPTEMBER FIFTEENTH. ALL VISITS BEGIN AT THE SURVEYOR'S HOUSE, THREE BLOCKS EAST OF THE CITY LIBRARY. ADMISSION INCLUDES A GUIDED TOUR OF THE SURVEYOR'S HOUSE AND INGALLS HOME, PLUS A TOURING MAP OF 16 OTHER SITES MENTIONED IN LAURA'S BOOKS.

SOUVENIRS AVAILABLE INCLUDE ALL HER BOOKS IN BOTH HARD COVER AND PAPERBACK, REPLICAS OF LAURA'S DOLL, COMMEMORATIVE PLATES, AND SUNBONNETS WITH APRONS TO MATCH.

A GRAND TOUR OF "LITTLE HOUSE" MEMORIALS WOULD INCLUDE RESTORATIONS, RECONSTRUCTIONS AND MARKED PLACES IN SEVERAL STATES. THE FIRST EIGHT BOOKS, THE "LITTLE HOUSE" CORE, ARE SHELVED INACCURATELY AS "FICTION" IN MOST LIBRARIES, BUT THEY ARE MEMOIRS WRITTEN IN THE THIRD PERSON. IT IS THE AUTHENTICITY OF PEOPLE, PLACE AND TIME THAT HAVE MADE THEM CONVINCING. IT IS THE REALITY THAT TAKES READERS OF ALL AGES TO THE VILLAGES ON THE PRAIRIE AND INTO THE WOODS OF THE OLD WEST. ALTHOUGH SEQUENCE IS NOT NECESSARY FOR

MICROFILMED FROM
BEST AVAILABLE COPY

ENJOYMENT, THE FIRST MEMORIAL OF THE GRAND TOUR WOULD BE THE WAYSIDE AT LAURA'S BIRTHPLACE NEAR PEPIN, WISCONSIN, DEDICATED IN 1979.

IN INDEPENDENCE, KANSAS, SITE OF THE ORIGINAL "LITTLE HOUSE ON THE PRAIRIE," A LOG CABIN FURNISHED WITH LOG FURNITURE OCCUPIES THE PROPERTY. ALAS, IT WAS BUILT AFTER THE INGALLS' TIME, BUT VISITORS MAY SEE THE CREEK AND BLUFFS MENTIONED IN THE STORY AND FILL IN THE BLANKS.

PA WILDER WAS A RESTLESS AND WANDERING MAN. THE NEXT HOME, WALNUT GROVE, MINNESOTA, SCENE OF "ON THE BANKS OF PLUM CREEK," HAS MARKED THE STAY WITH ITS OWN PAGEANT IN THE SCHOOL AUDITORIUM. "FRAGMENTS OF A DREAM" SETS WERE REPRODUCED FROM OLD PHOTOGRAPHS.

LAURA INGALLS WILDER, HER HUSBAND, ALMANZO, AND THEIR DAUGHTER, ROSE, WENT TO MISSOURI IN 1894 BY COVERED WAGON. THE FARMHOUSE AT ROCKY RIDGE IS WHERE THE BOOKS WERE WRITTEN, AND IS OPEN TO THE PUBLIC UNTIL THE SEASON ENDS OCTOBER FIFTEENTH.

A MUSEUM-GIFT SHOP IS ADJACENT TO THE HOME. THE HOUSE IS FULL OF MEMORABILIA; MARY'S ORGAN; PA'S FIDDLE; THE LAP DESK WHERE THE $100 BILL SO VITAL TO THEIR FUTURE WAS HIDDEN. THERE, TOO, ARE THE MANUSCRIPTS, WRITTEN

FIGURE D-1. (CONTINUED)

ON LINED YELLOW TABLETS FROM THE DIME STORE.

"IS IT TRUE?" MY DAUGHTER USED TO ASK WHEN SHE PUT DOWN A BOOK, "DID IT REALLY HAPPEN?"

IT IS, AND IT DID. THE PROOF IS WAITING ON THE PRAIRIE. YOU CAN STILL SEE THE COTTONWOOD TREES PA PLANTED, GROWING TALL.

THE BEST PLACE FOR PRE-TOUR HOMEWORK IS THE LAURA INGALLS WILDER ROOM OF THE POMONA PUBLIC LIBRARY, DEDICATED IN 1950. AT THAT TIME LAURA SENT A MARVELOUS GIFT, THE PENCILED MANUSCRIPT OF "LITTLE TOWN ON THE PRAIRIE." THIS TREASURE HAS BEEN ADDED TO, UNTIL POMONA IS NOW THE WEST COAST'S LARGEST REPOSITORY OF WILDER DATA AND RELICS. IT ALSO HAS CHARACTER FIGURES OF THE INGALLS FAMILY AND A WALL-SIZE ROUTE MAP OF THE INGALLS-WILDER TRAVELS.

-5-

MICROFILMED FROM
BEST AVAILABLE COPY

FIGURE D-1. (CONCLUDED)

YOU CAN TRAVEL IN CANADA WITHOUT PAYING FULL RATES FOR HOTELS AND MOTELS THIS SUMMER BY LOOKING INTO SOME OF THE ALTERNATIVES.

THEY RANGE FROM UNIVERSITY RESIDENCES AND COUNTRY HOMES TO THE OUTDOOR LIFE OF CAMPING, AND SPECIAL ACCOMMODATIONS DISCOUNTS AVAILABLE TO YOUTH AND STUDENTS. HERE ARE SOME OF THE DEALS AVAILABLE:

ANYONE OF ANY AGE CAN SEEK ACCOMMODATIONS IN 18 UNIVERSITY RESIDENCES ACROSS CANADA AT AN AVERAGE RATE OF $10 A NIGHT THIS SUMMER. THIS NEW VENTUREX PROGRAM IS CALLED THE AIR CANADA UNIPASS. IT'S A BOOKLET OF SEVEN VOUCHERS SOLD THROUGH TRAVEL AGENTS FOR $72. EACH VOUCHER IS EXCHANGEABLE FOR ONE NIGHT'S ACCOMMODATION (NO PRIOR BOOKING NECESSARY).

WITH THE VOUCHERS YOU WILL RECEIVE AN INFORMATION GUIDE THAT COVERS ADDRESSES, PHONE NUMBERS AND FACILITIES SUCH AS SWIMMING POOLS, SQUASH COURTS, TENNIS COURTS AND CAFETERIAS. THE GUIDE ALSO INDICATES THE COST OF REACHING A RESIDENCE FROM LOCAL AIR, BUS AND TRAIN TERMINALS.

IT'S A GOOD CONCEPT, BUT THERE ARE SOME IMPORTANT POINTS TO KEEP IN MIND WITH A PROGRAM OF THIS NATURE.

ALTHOUGH PRIOR BOOKING IS NOT NECESSARY, VOUCHERS DO NOT GUARANTEE THAT THERE WILL BE A BED LEFT FOR YOU, SO IT'S IN YOUR BEST INTEREST TO CALL AHEAD AND CHECK.

ALTHOUGH THE GUIDE INDICATES THAT ACCOMMODATIONS ARE AVAILABLE IN SEPARATE ROOMS, MANY UNIVERSITY ROOMS ARE DOUBLE, SO IF YOU ARRIVE BY YOURSELF IN A RUSH SEASON YOU COULD FIND YOURSELF WITH A ROOMMATE.

THERE ARE SEVEN VOUCHERS TO A BOOK, AND ONLY FULL BOOKS ARE REFUNDABLE, SO IF YOU DON'T USE ALL YOUR VOUCHERS YOU COULD BE OUT OF POCKET (UNLESS, OF COURSE, YOU CAN FIND SOMEONE ELSE WILLING TO BUY WHAT YOU HAVE LEFT). ANOTHER IMPORTANT POINT TO KEEP IN MIND IS LOCATION. UNIVERSITY RESIDENCES ARE IN BOTH CITIES AND SUBURBS. ALTHOUGH SUBURBAN ONES GENERALLY ARE ACCESSIBLE BY PUBLIC TRANSPORTATION, IT CAN COST COMMUTING TIME. ANOTHER LUXURY YOU MAY HAVE TO GIVE UP IN RESIDENCES IS A PRIVATE BATHROOM.

MOST OF THE UNIVERSITIES OPEN THEIR ACCOMMODATIONS TO VISITORS IN EARLY MAY AND CLOSE AT THE END OF AUGUST OR FIRST WEEK IN SEPTEMBER.

IF YOU'RE MORE THE TYPE TO ENJOY GETTING AWAY FROM THE CITIES AND ENJOYING THE COUNTRYSIDE, PRIVATE

FIGURE D-2. "CANADA" TEXT READ BY TEST SUBJECTS

MICROFILMED FROM BEST AVAILABLE COPY

HOMES TAKE IN GUESTS AT AN AVERAGE COST OF $15 A NIGHT SINGLE AND $18 DOUBLE. AN EXCELLENT SOURCE FOR FINDING THESE ESTABLISHMENTS IS JOHN THOMPSON'S "COUNTRY BED AND BREAKFAST PLACES IN CANADA."

IT'S A PAPERBACK LISTING OF MORE THAN 280 HOMES, THEIR LOCATIONS, PHONE NUMBERS AND SHORT DESCRIPTIONS WRITTEN BY EACH HOST FAMILY. AGAIN, IF YOU ARE GOING TO RELY ON THIS TYPE OF ACCOMMODATION, IT'S WISE TO CALL AHEAD AND CONFIRM THAT SPACE IS AVAILABLE.

RATES CAN BE AS LOW AS $12 DOUBLE IN SOME INSTANCES. ALTHOUGH THE'RE IN THE COUNTRY, THE LARGEST CONCENTRATION OF HOMES IS IN QUEBEC AND ONTARIO.

IF YOU ARE THE OUTDOOR TYPE CONSIDER THE POSSIBILITY OF CAMPING IN PRIVATE, PROVINCIAL AND FEDERAL CAMPGROUNDS. A NEW PUBLICATION AVAILABLE THIS SPRING OUTLINES THE FEDERAL CAMPING FACILITIES AND DESCRIBES EACH OF CANADA'S 28 NATIONAL PARKS. IT'S CALLED "NATIONAL PARKS -- A BRIEF GUIDE." IT'S NOT NECESSARY TO MAKE RESERVATIONS FOR CAMPSITES IN THE NATIONAL PARKS, BECAUSE THEY'RE AVAILABLE ON A FIRST-COME BASIS AND THERE'S A TWO-WEEK LIMIT FOR STAYING AT SITES. A PRIMITIVE, UNSERVICED SITE RUNS $5 A DAY; FULL SERVICE INCLUDING ELECTRICITY AND WATER IS $8.

FOR YOUNG TRAVELERS TRYING TO KEEP ACCOMMODATION COSTS TO A MINIMUM, MORE THAN 67 YOUTH HOSTELS WILL BE OPEN THIS SUMMER WITH RATES RUNNING $2 TO $8 A NIGHT. THIS DORMITORY STYLE OF LODGING (WITH SEPARATE ROOMS FOR MEN AND WOMEN) WILL BE AVAILABLE IN MANY CITIES AS WELL AS THE COUNTRYSIDE.

SOME OF THE LOCATIONS INCLUDE EIGHT HOSTELS IN BANFF NATIONAL PARK, NIAGARA FALLS, HISTORIC QUEBEC CITY, A BUILDING THAT WAS AT ONE TIME A JAIL IN OTTAWA AND EVEN A LOG CABIN IN MINTO FOR THOSE CANOEING ON THE YUKON RIVER.

THE HANDBOOK ALSO WILL GIVE YOU INFORMATION ABOUT OTHER TRAVEL DISCOUNTS AVAILABLE TO MEMBERS OF THE INTERNATIONAL YOUTH HOSTEL FEDERATION AND MENTION OF THE FEDERATION OF INTERNATIONAL YOUTH TRAVEL ORGANIZATIONS CARD. THE 70 YOUTH HOSTEL DISCOUNTS INCLUDE CYCLE RENTALS, ROCKY MOUNTAIN CYCLE TOURS, MUSEUM DISCOUNTS AND THE PROCEDURE FOR OBTAINING A SPECIAL STICKER THAT ENTITLES CARDHOLDERS TO CORPORATE RATES AT BUDGET CAR RENTAL OUTLETS.

ONE OF THE BEST BREAKS YOUTH TRAVEL CARD HOLDERS CAN GET IS A 50% DISCOUNT ON AVAILABLE ROOMS AT LUXURY CANADIAN HOTELS. THE ONLY REQUISITE FOR BEING A MEMBER IS THAT YOU BE UNDER 26. YOU SHOULD ASK FOR A CONCESSION

MICROFILMED FROM
BEST AVAILABLE COPY

FIGURE D-2. (CONTINUED)

LIST IF YOU DECIDE TO BUY THE CARD.

STUDENTS ALSO CAN GET TRAVEL DISCOUNTS ACROSS CANADA WHEN THEY HAVE AN INTERNATIONAL STUDENT IDENTITY CARD. THE CONCESSIONS INCLUDE CAR RENTALS, RIVER, CYCLE AND HORSEBACK TOURS, CAMPING SITES, HOTELS AND MOTELS, RESTAURANTS AND STUDENT RESIDENCE SUMMER ACCOMMODATIONS. ONE OF THE BEST DEALS IS THAT SOME OF THE SHERATON INNS AND HOTELS WILL OFFER 25% REDUCTIONS WITH A RIGHT TO REMOVE THEM AT PEAK OCCUPANCY PERIODS.

-5-

MICROFILMED FROM
BEST AVAILABLE COPY

FIGURE D-2. (CONCLUDED)

MICROFILMED FROM
BEST AVAILABLE COPY

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-3547		

4. Title and Subtitle	5. Report Date
MEASUREMENT OF SPEECH LEVELS IN THE PRESENCE OF TIME VARYING BACKGROUND NOISE	April 1982
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Karl S. Pearsons and Richard D. Horonjeff	4554
	10. Work Unit No.

9. Performing Organization Name and Address	11. Contract or Grant No.
Bolt Beranek and Newman Inc. 21120 Vanowen Street Canoga Park, California 91305	NAS1-16521
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Contractor Report
National Aeronautics and Space Administration Washington, D. C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

Langley technical monitor: Clemans A. Powell, Jr.
Final Report

16. Abstract

A study was performed to develop short-term speech level measurements which could be used to note changes in vocal effort in a time-varying noise environment. Knowing the changes in speech level would in turn allow prediction of intelligibility in the presence of aircraft flyover noise. Tests indicated that it is possible to use 2 second samples of speech to estimate long-term rms speech levels. Other tests were also performed in which people read out loud during aircraft flyover noise. Results of these tests indicate that people do indeed raise their voice during flyovers at a rate of about 3-1/2 dB for each 10 dB increase in background level. This finding is in agreement with other tests of speech levels in the presence of steady-state background noise.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Speech measurement Aircraft noise Time-varying noise	Unclassified-Unlimited Subject Category 45

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	76	A05

For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1982